

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:17:56 ; Search time 48.03 Seconds
(without alignments)
607.430 Million cell updates/sec

Title: US-09-072-384-18
Perfect score: 2080
Sequence: 1 MAGIPGLLFLFLCAVQ.....LKVAQICYWKNYLDREG 383

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	194	9.3	316	2 A45134	endopeptidase (EC 3.4.21.1)
2	135.5	6.5	313	2 A35122	metalloproteinase
3	120.5	5.8	482	1 EXPT	coagulation factor
4	119	5.7	269	2 A26823	pancreatic elastase
5	117	5.6	522	2 T29767	hypothetical prote
6	115.5	5.6	269	2 B26823	pancreatic elastase
7	115.5	5.6	271	2 A25528	pancreatic elastase
8	115	5.5	258	2 S70439	pancreatic elastase
9	115	5.5	267	4 A56615	probable pancreati
10	114.5	5.5	238	1 TRHV5Y	trypsin-like prote
11	114	5.5	246	1 DBHU	complement factor
12	113	5.4	266	1 ELRT1	pancreatic elastase
13	113	5.4	266	1 ELPG	pancreatic elastase
14	112.5	5.4	259	1 TRSMG	trypsin (EC 3.4.21
15	112.5	5.4	273	2 E85765	hypothetical prote
16	112.5	5.4	273	2 H64915	hypothetical prote
17	111.5	5.4	488	1 EXHU	coagulation factor
18	111	5.3	761	2 JC5759	brain-specific ser
19	110	5.3	269	2 C26823	pancreatic elastase
20	110	5.3	492	1 EXBO	coagulation factor
21	110	5.3	1582	2 T15308	hypothetical prote
22	109.5	5.3	405	2 T35117	probable secreted
23	108	5.2	583	2 A25154	complement factor
24	108	5.2	786	1 A47547	serine proteinase
25	108	5.2	1047	2 A55617	masquerade precurs
26	107.5	5.2	236	2 A28566	T-cell suppressor
27	107.5	5.2	686	1 A59271	Ra-reactive factor
28	105	5.0	274	2 SA0004	trypsin-related pr
29	103.5	5.0	271	1 ELRT2	pancreatic elastase

30	103.5	5.0	416	1 KFBO	coagulation factor
31	102.5	4.9	1238	2 T34929	hypothetical prote
32	100	4.8	272	2 JC4170	trypsin-like prote
33	98.5	4.7	267	2 S40006	trypsin (EC 3.4.21
34	98	4.7	548	2 D82175	probable trypsin V
35	98	4.7	624	2 T02289	probable polyalac
36	97.5	4.7	409	2 T35118	probable secreted
37	97.5	4.7	452	1 A30351	coagulation factor
38	97.5	4.7	747	2 I51579	complement factor
39	97	4.7	1019	2 A38738	coagulation factor
40	96.5	4.6	259	2 S68424	allergen Der f III
41	96.5	4.6	275	2 I46712	factor IX - rabbit
42	96.5	4.6	2145	2 JC4747	adenylate cyclase
43	95.5	4.6	430	1 A24702	serine proteinase
44	95	4.6	782	2 T32155	hypothetical prote
45	94	4.5	268	2 S68825	pancreatic elastase

ALIGNMENTS

RESULT 1
A45134
endopeptidase (EC 3.4.21.1), glutamate-specific - Bacillus licheniformis
C:Species: Bacillus licheniformis
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
C:Accession: A45134; S23078
R:Kakudo, S.; Kikuchi, N.; Kitadokoro, K.; Fujiwara, T.; Nakamura, E.; Okamoto, H.; S
J. Biol. Chem. 267, 23782-23788, 1992
A:Title: Purification, characterization, cloning, and expression of a glutamic acid-s
A:Reference number: A45134; MUID:93054737
A:Accession: A45134
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-316 <KAK>
A:Cross-References: GB:D10060; NID:9216263; PIDN:BAA00949.1; PID:d1001415; PID:g21626
A:Experimental source: ATCC 14580
A:Note: sequence extracted from NCBI backbone (NCBIN:118784, NCBIP:118785)
R:Svendsen, I.; Bredem, K.
Eur. J. Biochem. 204, 165-171, 1992
A:Title: Isolation and amino acid sequence of a glutamic acid specific endopeptidase
A:Reference number: S23078; MUID:92155199
A:Accession: S23078
A:Status: preliminary
A:Molecule type: protein
A:Residues: 95-316 <SVR>
C:Keywords: hydrolase

Query Match 9.3%; Score 194; DB 2; Length 316;

Best Local Similarity 24.0%; Pred. No. 1.2e-08;

Matches 79; Conservative 43; Mismatches 143; Indels 64; Gaps 15;

QY	63	SSCGPOCHKGTPL---PTVEAKQYLSYETLYANGSRTETQVGIYLLSSSGCAQHRDSC	119
DB	28	AQAAPSPH--TPVSDPSY-KAETSVDYD-----NIKSDQYGLYKAFGTGKVNETKE	79
QY	120	SSGSKRRKQIYGYDGRSTFGKDFLLN-----YFESTSVKLSL---GCTGLVAEKHV	170
DB	80	KAEKSPAKAPY---SIKSVIGSDDRVTNTTAYPRATVHSSISGSGCTGMIGPKTV	136
QY	171	LTAACHIDCKT-YVKGTOKLRVGLKPKFKDGGRCANDSTAMPQMKFQWIRVRKTHV	229
DB	137	ATAGHCIDYTSSTSGTAGTATVSPG-----RNCTS-----YPYGVSKSTRYFI	178
QY	230	PKGWIKGNANDICMDYDYLLELKKPKHKFKMIGVSPPAKQLPGGRHFSGYDNDRPGN	289
DB	179	PSGWSGNTN-----YDYGAIELSEPIGNTVGVFGYSYTTSSLVGTTVTSYGGPKTAG	233
QY	290	LVYRECD--VKDETLLYQOCDAQPGASGVVVRMKRQOKWERKII---GIFSGH	343
DB	234	TQWHSGLPAISYTKLQYAM-DTYGQSGSPVFESSSRNTCSGPCSLAVHTNGVYGG-	291

Qy 344 QWDMNGSPQDENVAVRITPLKVAQICW 372
Db 292 -----SSYNGRTTRITKEVFDNLTNW 311

RESULT 2
Metalloproteinase (EC 3.4.21.6) mpr precursor, extracellular - Bacillus subtilis
A:Species: Bacillus subtilis
C:Date: 27-Jul-1990 #sequence_revision 27-Jul-1990 #text_change 15-Oct-1999
C:Accession: A35122; I40010; A69660
R:Sloma, A.; Rudolph, C.F.; Rufo Jr., G.A.; Sullivan, B.J.; Theriault, K.A.; Ally, D.; P.
J. Bacteriol. 172, 1024-1029, 1990
A:Title: Gene encoding a novel extracellular metalloprotease in Bacillus subtilis.
A:Reference number: A35122; MUID:90130256
A:Accession: A35122
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-313 <SLO>
A:Cross-references: GB:L10505; NID:g143209; PIDN:AAA22604.1; PID:g143210; GB:M29036
R:Smith, H.; de Jong, A.; Bron, S.; Venema, G.
Gene 70, 351-361, 1988
A:Title: Characterization of signal-sequence-coding regions selected from the Bacillus s
A:Reference number: I39994; MUID:89108019
A:Accession: I40010
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-60,65,'L',67,'S',69,'AOA' <RES>
A:Cross-references: GB:M2916; NID:g143701; PIDN:AAA22832.1; PID:g143702
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berten
C.; Bron, S.; Brouillet, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
A:Authors: Schleich, S.; Schroeder, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033
A:Accession: A69660
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-313 <KUN>
A:Cross-references: GB:299105; GB:AL009126; NID:g2632457; PIDN:CAB12018.1; PID:ell82176;
A:Experimental source: strain 168
C:Genetics:
A:Gene: mpr
C:Keywords: hydrolase

Query Match 6.5%; Score 135.5; DB 2; Length 313;
Best Local Similarity 22.2%; Pred. No. 0.00091;
Matches 77; Conservative 38; Mismatches 121; Indels 111; Gaps 17;

Qy 72 GTPLPTVEAKQYLSYETLYANGSRTETQVGIYLLSSGD--GAQRDSCGSKSRKRQ 129
Db 29 GVPAAKAENPQTSVNTGKADATKNOT-----SKADQVSAPYEGTGKTSK----- 75

Qy 130 IYGVDSRF-----SIFGKD-----FLNLPSTSVKLEST-----GCTGTL 164
Db 76 LYGGQTELEKNIQPLSPSSIIIGTDERTRISTTSFPRAIVQLSIKYPNVSSTYGCCTGFL 135

Qy 165 VAEKHVLTAACHIH-----DGKTYVKGQKLRGFLKPKFKDGGRGAND 208
Db 136 VNPNTVTVAGHCYVSDHQWASTITAAAPGRNGSSYPYGVTV----- 175

Qy 209 STSAMPQMKFQWTRVTRVTHVPGKWIKN---ANDICMDYDYLLELKKPKRKMKGIV 265

Db 176 -SGTMFYSVK-GWTESKDTNDYGAIKLNGSPGNTVGW-YGYRTTNSSP-----VGL 225
Qy 266 SPPAKQLPGGRHFGSYNDPRGNLVYFCDKDYDLYLLYQCDQAQSGSCSGYVYRMW 325
Db 226 SSSVTGFPCKDTFGTMWSDTKPIR-----SAETYLKLT-TTDTYGCQSGSPVY----- 272

Qy 326 KRQCKWKERKIIGFSGHQWDMNGSPQDENVAVRITPLKVAQICW 372
Db 273 -RNYSDTGQTALHT-----NGG-SSYNLGTRVNDVFNNTQYW 310

RESULT 3
EXRT
coagulation factor Xa (EC 3.4.21.6) precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Jan-1995 #sequence_revision 07-Feb-1997 #text_change 08-Dec-2000
C:Accession: S49075; JC4670; PS0191; PS0190; I62745
R:Stanton, C.; Ross, P.; Hutson, S.; Wallin, R.
Thromb. Res. 80, 63-73, 1995
A:Title: Evidence for competition between vitamin K-dependent clotting factors
A:Reference number: A58498; MUID:96093366
A:Accession: S49075
A:Molecule type: mRNA
A:Residues: 1-482 <STAL>
A:Cross-references: EMBL:X79807; NID:g506600; PIDN:CAA56202.1; PID:g506601
A:Note: submitted to the EMBL Data Library, June 1994
A:Note: neither the complete nucleic acid sequence nor the complete translation are s
R:Stanton, C.; Ross, R.P.; Hutson, S.; Wallin, R.
Gene 189, 269-273, 1996
A:Title: Processing and expression of rat and human clotting factor-X-encoding cDNAs.
A:Reference number: JC4670; MUID:96194815
A:Accession: JC4670
A:Molecule type: mRNA
A:Residues: 1-482 <STAP2>
A:Cross-references: EMBL:X79807; NID:g506600; PIDN:CAA56202.1; PID:g506601
A:Experimental source: Cos-1 cell
R:Enyoji, K.; Miyazaki, K.; Kato, H.
J. Biochem. 109, 890-898, 1991
A:Title: Characterization of rat factors X and Xa: demonstration of factor Xa in rat
A:Reference number: PS0190; MUID:92041742
A:Accession: PS0191
A:Molecule type: protein
A:Residues: 41-58,'X',60-65 <ENJ1>
A:Accession: PS0190
A:Molecule type: protein
A:Residues: 183-186,'X',188-207 <ENJ2>
R:Murakawa, M.; Okamura, T.; Kamura, T.; Kuroiwa, M.; Harada, M.; Niho, Y.
Eur. J. Haematol. 52, 162-168, 1994
A:Title: Analysis of the partial nucleotide sequences and deduced primary struct
A:Reference number: I46196; MUID:94222160
A:Accession: I62745
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Residues: 295-383,'G',385-455 <MUR>
A:Cross-references: GB:D21215; NID:g415309; PIDN:BAA04756.1; PID:g455396
C:Function:
A:Description: catalyzes the proteolytic activation of prothrombin to thrombin in the
A:Pathway: blood coagulation
C:Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homol
C:Keywords: beta-hydroxyaspartic acid; blood coagulation; calcium binding; carboxyglu
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-40/Domain: propeptide #status predicted <PRO>
F:25-84/Domain: Gla domain homology <GLA>
F:41-179/Product: coagulation factor X light chain #status predicted <LCH>
F:90-121/Domain: EGF homology <EG1>
F:129-164/Domain: EGF homology <EG2>
F:183-482/Product: coagulation factor X heavy chain #status predicted <HCH>
F:183-231/Domain: activation peptide #status predicted <APT>
F:232-482/Product: coagulation factor Xa heavy chain #status predicted <ACT>
F:232-460/Domain: trypsin homology <TRY>
F:45-47,54,56,59,60,65,66,69,72,79/Modified site: gamma-carboxyglutamic acid (Glu) #s
F:57-62,90-101,95-110,112-121,129-140,136-149,151-164,172-340,238-243,259-275,388-402

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:25:48 ; Search time 84.84 Seconds
(without alignments)
675.846 Million cell updates/sec

Title: US-09-072-384-15
Perfect score: 2131
Sequence: 1 MAGIPGLFLFLFLCAVGG.....IKGYLDCREGDTVFPPGSN 392

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	DB ID	Description
1	2080	97.6	383	4	Q95084
2	1884.5	88.4	382	11	Q9D6X6
3	1042	48.9	413	4	Q9BQP6
4	133.5	6.3	799	11	Q9DBI0
5	129.5	6.1	303	2	Q9EXR9
6	129	6.1	469	6	Q9GMD9
7	127.5	6.0	1322	5	Q9NAT0
8	126.5	5.9	678	11	Q9JJJ58
9	124.5	5.8	1322	5	Q9NJS5
10	124	5.8	266	6	O46644
11	123.5	5.8	339	11	Q9QX91
12	123.5	5.8	366	11	Q9QX85
13	123.5	5.8	541	11	Q9QX90
14	123.5	5.8	623	11	Q9JJJ3
15	123.5	5.8	643	11	Q9QX84
16	122	5.7	259	5	Q9XY61
17	122	5.7	449	5	Q9YD08
18	120.5	5.7	482	11	O63207
19	118	5.5	266	11	Q9D936

Q9V516 drosophila
Q9VQR8 drosophila
Q9W5U8 drosophila
O01771 caenorhabdi
Q14243 homo sapien
Q9UN11 homo sapien
Q9G991 manduca sex
Q9VA87 drosophila
Q9Z338 mus musculu
Q90422 brachydanio
Q9VDV1 drosophila
Q9BZH0 homo sapien
Q9W7Q1 paralichthy
Q54740 mus musculu
Q10922 caenorhabdi
Q69973 streptomyce
Q88947 mus musculu
Q99132 mus musculu
Q17086 anopheles s
Q9W633 cyprinus ca
Q24019 drosophila
Q9ZLH1 mus musculu
Q9AYR4 chaetoceros
Q9VZS8 drosophila
Q9Vfz6 drosophila
Q9Yic6 cyprinus ca

ALIGNMENTS

RESULT 1
O95084
ID O95084 PRELIMINARY; PRT; 383 AA.
AC O95084;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SERINE PROTEASE (HYPOTHETICAL 43.0 KDA PROTEIN) (PROTEASE, SERINE,
DE 23).
GN ZSGL3 OR DKFZP586B0719.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-UMBILICAL VEIN;
RA Li X., Tedder T.F.;
RT "A novel serine protease from human umbilical vein endothelial
RT cells";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Sheppard P., Blumberg H., Jelinek L., Foster D., O'Hara P.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-UTERUS;
RA Wambutt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP TISSUE-CERVIX CARCINOMA;
RA Strausberg R.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC -!- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
DR EMBL: AF015287; AAD01553.1;
DR EMBL: AF193611; AAF07186.1;
DR EMBL: AF136914; CAB86848.1;
DR EMBL: BC001278; AAH01278.1;
DR MEROPS; S01.309; -.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:16:53 ; Search time 64.47 Seconds
(without alignments)
450.391 Million cell updates/sec

Title: US-09-072-384-15
Perfect score: 2131
Sequence: 1 MAGIPGLFLFLFLLCAVCG.....IKGNLYDCRGDTVPFPGSN 392

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2131	100.0	392	22	Human Zsig13 varia
2	2085	97.8	392	22	Human Zsig13 varia
3	2080	97.6	383	20	Human transmembran
4	2080	97.6	383	20	WC9927094 Seq ID 1
5	2080	97.6	383	20	Amino acid sequenc
6	2080	97.6	383	21	Protein encoded by
7	2080	97.6	383	21	Protein encoded by
8	2080	97.6	383	21	Human TANGO 186 pr
9	2080	97.6	383	21	Human signal pepti
10	2080	97.6	383	21	A bone marrow secr
11	2080	97.6	383	22	Human PRO307 prote

12	2080	97.6	383	22	Human Zsig13 varia
13	2073	97.3	383	22	Human membrane or
14	1909.5	89.6	375	21	Protein encoded by
15	1909.5	89.6	413	21	Protein encoded by
16	1890.5	88.7	382	21	Murine TANGO 186 p
17	1041	48.9	413	20	Human PRO1057 prot
18	1041	48.9	413	20	Human nn320_2 secr
19	1041	48.9	413	21	Human PRO1057 (UNQ
20	202.5	9.5	314	22	Bacillus lichenifo
21	194	9.1	316	13	Bacillus lichenifo
22	193	9.1	316	22	Protease BLase. S
23	184	8.6	222	22	Bacillus lichenifo
24	183	8.6	222	22	Bacillus lichenifo
25	183	8.6	222	22	Bacillus lichenifo
26	183	8.6	222	22	Bacillus lichenifo
27	183	8.6	222	22	Bacillus lichenifo
28	183	8.6	222	22	Bacillus lichenifo
29	181	8.5	222	22	Bacillus lichenifo
30	181	8.5	222	22	Bacillus lichenifo
31	181	8.5	222	22	Bacillus lichenifo
32	181	8.5	222	22	Bacillus lichenifo
33	180	8.4	222	22	Bacillus lichenifo
34	180	8.4	222	22	Bacillus lichenifo
35	180	8.4	222	22	Bacillus lichenifo
36	180	8.4	222	22	Bacillus lichenifo
37	179	8.4	222	22	Bacillus lichenifo
38	179	8.4	222	22	Bacillus lichenifo
39	179	8.4	222	22	Bacillus lichenifo
40	179	8.4	222	22	Bacillus lichenifo
41	179	8.4	222	22	Bacillus lichenifo
42	179	8.4	222	22	Bacillus lichenifo
43	179	8.4	222	22	Bacillus lichenifo
44	179	8.4	222	22	Bacillus lichenifo
45	179	8.4	222	22	Bacillus lichenifo

ALIGNMENTS

RESULT 1	
AA848973	
ID	AA848973 standard; Protein; 392 AA.
AC	AA848973;
XX	
DT	27-MAR-2001 (first entry)
XX	
DE	Human Zsig13 variant #2, SEQ ID NO:15.
XX	
KW	Human Zsig13; serine protease; chromosome 11q22.1; elastase homologue;
KW	glutaryl endopeptidase homologue; factor X homologue; trypsin homologue;
KW	trypsinogen homologue; mast cell protease homologue;
KW	collagenase homologue; protein degradation; food processing; brewing;
XX	alcohol production; laundry detergent component.
OS	Homo sapiens.
XX	
PN	US6153420-A.
XX	
PD	28-NOV-2000.
XX	
PF	04-MAY-1998; 98US-0072384.
XX	
PR	24-APR-1997; 97US-0044185.
PR	17-APR-1998; 98US-0062142.
XX	
PA	(ZYMO) ZYMOGENETICS INC.
XX	
PI	Sheppard PO;
XX	
DR	WPI; 2001-060090/07.
XX	
DR	N-PSDB; AAC91783.
XX	

PT New isolated serine protease (designated zsig13), useful in industrial
PT processes to degrade unwanted proteins or alter the characteristics of
PT protein-containing composition, as well as in industrial applications
PT (e.g. brewing)
XX

PS Claim 1; Column 35-38; 26pp; English.

XX The invention relates to human zsig13 proteins (AAB48972-B48974), and
CC to DNA encoding them (AAC91782-C91784). The invention also relates to
CC expression vectors and host cells comprising a human zsig13 DNA, and the
CC recombinant production of a human zsig13 protein or its precursor.
CC zsig13 is a serine protease, and has significant homology to Bacillus
CC licheniformis glutamyl endopeptidase, human clotting factor X, human
CC elastase, rat mast cell protease, Streptomyces griseus trypsin, bovine
CC trypsinogen, and Hypoderma lineatum collagenase. The gene encoding human
CC zsig13 is located on chromosome 11q22.1. zsig13 is useful in industrial
CC processes to degrade unwanted proteins or alter the characteristics of
CC protein-containing compositions. It may also be used in industrial
CC applications in which proteases are utilised, including food processing,
CC brewing and alcohol production, and as a component of a laundry
CC detergent. The present sequence represents a human zsig13 variant.
XX

SQ Sequence 392 AA;

Query Match 100.0%; Score 2131; DB 22; Length 392;
Best Local Similarity 100.0%; Pred. No. 3.8e-152;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGIPGLLLFLLLCAGGVSPYSAPWKTWPAYRLPVVLPDSTLNLAKEPFGAEAKLE 60

DB 1 magipglilflilfcavgvspysapwktwpayrlpvvlpdstlnlakpdfgaeakle 60

QY 61 VSSSCGQCHKGTPLYTEAKOYLSVETLYANGSRTETOVGIYILSSSGDGAQHRDGS 120

DB 61 vssscgqchkgtplyteakylsyetlyangsrteetovgiylsssgdgqhrds 120

QY 121 SGKSRKRQIYGYDSRFSIFGKDFLLNPFSTSVKLTGCTGLTVAEKHVLTAACHIDG 180

DB 121 sgksrrkrqlygydsrfsifgkdfllnypfstsvklstgctglvaekhvltaachidg 180

QY 181 KTYVKGTKLRVGFLLPKFKDGGRGANDSTAMPQKMFQWIRVTRKTHVPKGWIKGNAND 240

DB 181 ktyvkgtklrvglfllpkfkdggrgandstampsqkmgfqrwtrkthvpgkwikgnand 240

QY 241 IGMDDYVALLLEKPKHKKRPMKIGVSPPAKQLPGGRHFSGYDNDRPNLVYRFGCDVKDE 300

DB 241 igmddyvallelkpkhkrkfmkigvspapakqlpggrihfsgydnndrpnlvyrfgcdvkde 300

QY 301 TYDLLYQQCDAQPGASGSYVYVVMWKRQOKWERKIIIGIFSGHQWDMNGSPQDFNVAVR 360

DB 301 tydillyqqcdagpgasgvyvymwkrqqkwerkiigifsghqwvdmngspqdfnavr 360

QY 361 ITPKVAQICYWIKGNVLDCREGDTVFPFGSN 392

DB 361 itplkvaqicywikgnvldcregdtvfpfgsn 392

RESULT 2

AAB48972

ID AAB48972 standard; Protein: 392 AA.

XX

AC AAB48972;

XX

DT 27-MAR-2001 (first entry)

DE Human zsig13 variant #1, SEQ ID NO:2.

XX

KW Human zsig13; serine protease; chromosome 11q22.1; elastase homologue;
KW glutamyl endopeptidase homologue; factor x homologue; trypsin homologue;
KW trypsinogen homologue; mast cell protease homologue;
KW collagenase homologue; protein degradation; food processing; brewing;
KW alcohol production; laundry detergent component.

XX Homo sapiens.

OS US6153420-A.

PN 28-NOV-2000.

XX

PD 04-MAY-1998; 98US-0072384.

XX

PF 24-APR-1997; 97US-0044185.

XX

PR 17-APR-1998; 98US-0062142.

XX

PA (ZYMO) ZYMOGENETICS INC.

XX

PI Sheppard PO.

XX

XX WPI; 2001-060090/07.

DR N-PSDB; AAC91782.

XX

PT New isolated serine protease (designated zsig13), useful in industrial

PT processes to degrade unwanted proteins or alter the characteristics of

PT protein-containing composition, as well as in industrial applications

PT (e.g. brewing)

XX

PS Claim 1; Column 25-28; 26pp; English.

XX

CC The invention relates to human zsig13 proteins (AAB48972-B48974), and

CC to DNA encoding them (AAC91782-C91784). The invention also relates to

CC expression vectors and host cells comprising a human zsig13 DNA, and the

CC recombinant production of a human zsig13 protein or its precursor.

CC zsig13 is a serine protease, and has significant homology to Bacillus

CC licheniformis glutamyl endopeptidase, human clotting factor X, human

CC elastase, rat mast cell protease, Streptomyces griseus trypsin, bovine

CC trypsinogen, and Hypoderma lineatum collagenase. The gene encoding human

CC zsig13 is located on chromosome 11q22.1. zsig13 is useful in industrial

CC processes to degrade unwanted proteins or alter the characteristics of

CC protein-containing compositions. It may also be used in industrial

CC applications in which proteases are utilised, including food processing,

CC brewing and alcohol production, and as a component of a laundry

CC detergent. The present sequence represents a human zsig13 variant.

XX

SQ Sequence 392 AA;

Query Match 97.8%; Score 2085; DB 22; Length 392;

Best Local Similarity 98.2%; Pred. No. 1.1e-148;

Matches 385; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MAGIPGLLLFLLLCAGGVSPYSAPWKTWPAYRLPVVLPDSTLNLAKEPFGAEAKLE 60

DB 1 magipglilflilfcavgvspysapwktwpayrlpvvlpdstlnlakpdfgaeakle 60

QY 61 VSSSCGQCHKGTPLYTEAKOYLSVETLYANGSRTETOVGIYILSSSGDGAQHRDGS 120

DB 61 vssscgqchkgtplyteakylsyetlyangsrteetovgiylsssgdgaxxrds 120

QY 121 SGKSRKRQIYGYDSRFSIFGKDFLLNPFSTSVKLTGCTGLTVAEKHVLTAACHIDG 180

DB 121 sgksrrkrqlygydsrfsifgkdfllnypfstsvklstgctglvaekhvltaachidg 180

QY 181 KTYVKGTKLRVGFLLPKFKDGGRGANDSTAMPQKMFQWIRVTRKTHVPKGWIKGNAND 240

DB 181 ktyvkgtklrvglfllpkfkdggrgandstampsqkmgfqrwtrkthvpgkwikgnand 240

QY 241 IGMDDYVALLLEKPKHKKRPMKIGVSPPAKQLPGGRHFSGYDNDRPNLVYRFGCDVKDE 300

DB 241 igmddyvallelkpkhkrkfmkigvspapakqlpggrihfsgydnndrpnlvyrfgcdvkde 300

QY 301 TYDLLYQQCDAQPGASGSYVYVVMWKRQOKWERKIIIGIFSGHQWDMNGSPQDFNVAVR 360

DB 301 tydillyqqcdagpgasgvyvymwkrqqkwerkiigifsghqwvdmngspqdfnavr 360

QY 361 ITPKVAQICYWIKGNVLDCREGDTVFPFGSN 392

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:15:34 ; Search time 35.78 Seconds
(without alignments)
240.882 Million cell updates/sec

Title: US-09-072-384-18
Perfect score: 2080
Sequence: 1 MAGIGLLFLFLFLCAVQ.....LKVAICYWIKGNVLDREG 383

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2080	100.0	383	4	US-09-072-384-18
2	2080	100.0	392	4	US-09-072-384-15
3	2044	98.3	392	4	US-09-072-384-2
4	177	8.5	222	1	US-08-090-048-1
5	177	8.5	222	2	US-08-292-550-1
6	177	8.5	222	2	US-07-927-661A-1
7	122	5.9	256	3	US-08-906-769-89
8	122	5.9	256	3	US-08-906-616-89
9	122	5.9	256	4	US-08-817-795-89
10	122	5.9	256	4	US-08-639-075A-89
11	122	5.9	256	4	US-09-012-431-89
12	122	5.9	256	4	US-09-032-215-32
13	122	5.9	256	4	US-09-012-693-89
14	122	5.9	256	4	US-08-906-613-89
15	122	5.9	256	5	PCT-US95-14442A-89
16	117	5.6	437	1	US-08-487-037-2
17	115.5	5.6	241	4	US-08-944-483-59
18	114	5.5	228	4	US-08-944-483-44
19	114	5.5	253	6	5223425-8
20	113	5.4	238	6	5223425-5
21	113	5.4	250	6	5223425-4
22	112.5	5.4	223	1	US-08-278-091-13
23	112.5	5.4	223	1	US-08-483-859-13
24	112.5	5.4	223	1	US-08-472-173-13
25	112.5	5.4	223	2	US-08-487-167-13
26	112.5	5.4	223	2	US-08-482-816-13
27	112.5	5.4	223	2	US-08-296-149-13

28	112.5	5.4	223	2	US-08-801-499-13	Sequence 13, Appl
29	112.5	5.4	223	3	US-08-615-271-13	Sequence 13, Appl
30	112.5	5.4	223	3	US-09-074-660-13	Sequence 13, Appl
31	112.5	5.4	223	3	US-09-074-659-13	Sequence 13, Appl
32	112.5	5.4	223	3	US-09-106-468-13	Sequence 13, Appl
33	112.5	5.4	223	4	US-09-106-468A-13	Sequence 13, Appl
34	112.5	5.4	223	4	US-09-106-467-13	Sequence 13, Appl
35	111.5	5.4	241	1	US-08-330-978-4	Sequence 4, Appl
36	111.5	5.4	241	1	US-08-474-042-4	Sequence 4, Appl
37	111.5	5.4	241	1	US-08-484-558-4	Sequence 4, Appl
38	111.5	5.4	241	1	US-08-774-592-4	Sequence 4, Appl
39	111.5	5.4	254	1	US-08-330-978-3	Sequence 3, Appl
40	111.5	5.4	254	1	US-08-474-042-3	Sequence 3, Appl
41	111.5	5.4	254	1	US-08-484-558-3	Sequence 3, Appl
42	111.5	5.4	254	1	US-08-774-592-3	Sequence 3, Appl
43	111.5	5.4	306	1	US-08-330-978-1	Sequence 1, Appl
44	111.5	5.4	306	1	US-08-474-042-1	Sequence 1, Appl
45	111.5	5.4	306	1	US-08-484-558-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-072-384-18
; Sequence 18, Application US/09072384
; Patent No. 6153420
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
; TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,384
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31,648
; REFERENCE/DOCKET NUMBER: 97-16C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 383 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
; FEATURE:
; NAME/KEY: Signal Sequence
; LOCATION: 1..19
; OTHER INFORMATION:
US-09-072-384-18

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:16:53 ; Search time 64.47 Seconds
(without alignments)
440.051 Million cell updates/sec

Title: US-09-072-384-18
Perfect score: 2080
Sequence: 1 MAGIPGLFLFLFLCAVCG.....LKQAICYWIKGNLYDCREG 383

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101:*

1: /SID52/gcgdata/geneseq/geneseq/AA1980.DAT:*

2: /SID52/gcgdata/geneseq/geneseq/AA1981.DAT:*

3: /SID52/gcgdata/geneseq/geneseq/AA1982.DAT:*

4: /SID52/gcgdata/geneseq/geneseq/AA1983.DAT:*

5: /SID52/gcgdata/geneseq/geneseq/AA1984.DAT:*

6: /SID52/gcgdata/geneseq/geneseq/AA1985.DAT:*

7: /SID52/gcgdata/geneseq/geneseq/AA1986.DAT:*

8: /SID52/gcgdata/geneseq/geneseq/AA1987.DAT:*

9: /SID52/gcgdata/geneseq/geneseq/AA1988.DAT:*

10: /SID52/gcgdata/geneseq/geneseq/AA1989.DAT:*

11: /SID52/gcgdata/geneseq/geneseq/AA1990.DAT:*

12: /SID52/gcgdata/geneseq/geneseq/AA1991.DAT:*

13: /SID52/gcgdata/geneseq/geneseq/AA1992.DAT:*

14: /SID52/gcgdata/geneseq/geneseq/AA1993.DAT:*

15: /SID52/gcgdata/geneseq/geneseq/AA1994.DAT:*

16: /SID52/gcgdata/geneseq/geneseq/AA1995.DAT:*

17: /SID52/gcgdata/geneseq/geneseq/AA1996.DAT:*

18: /SID52/gcgdata/geneseq/geneseq/AA1997.DAT:*

19: /SID52/gcgdata/geneseq/geneseq/AA1998.DAT:*

20: /SID52/gcgdata/geneseq/geneseq/AA1999.DAT:*

21: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT:*

22: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2080	100.0	383	20	Human transmembran
2	2080	100.0	383	20	WO9927094 Seq ID 1
3	2080	100.0	383	20	Amino acid sequenc
4	2080	100.0	383	21	Protein encoded by
5	2080	100.0	383	21	Protein encoded by
6	2080	100.0	383	21	Human TANGO 186 pr
7	2080	100.0	383	21	Human signal pepti
8	2080	100.0	383	21	A bone marrow secr
9	2080	100.0	383	22	Human PRO307 prote
10	2080	100.0	383	22	Human Zsig13 varia
11	2080	100.0	392	22	Human Zsig13 varia

12	2073	99.7	383	22	Human membrane or
13	2044	98.3	392	22	Human Zsig13 varia
14	1909.5	91.8	375	21	Protein encoded by
15	1909.5	91.8	413	21	Protein encoded by
16	1890.5	90.9	382	21	Murine TANGO 186 p
17	1041	50.0	413	20	Human PRO1057 prot
18	1041	50.0	413	21	Human nn320_2 secr
19	1041	50.0	413	20	Human PRO1057 (UNQ
20	202.5	9.7	314	22	Bacillus lichenifo
21	194	9.3	316	13	Protease Blase. S
22	193	9.3	316	22	Bacillus lichenifo
23	184	8.8	222	22	Bacillus lichenifo
24	183	8.8	222	22	Bacillus lichenifo
25	183	8.8	222	22	Bacillus lichenifo
26	183	8.8	222	22	Bacillus lichenifo
27	183	8.8	222	22	Bacillus lichenifo
28	183	8.8	222	22	Bacillus lichenifo
29	181	8.7	222	22	Bacillus lichenifo
30	181	8.7	222	22	Bacillus lichenifo
31	181	8.7	222	22	Bacillus lichenifo
32	181	8.7	222	22	Bacillus lichenifo
33	180	8.7	222	22	Bacillus lichenifo
34	180	8.7	222	22	Bacillus lichenifo
35	180	8.7	222	22	Bacillus lichenifo
36	180	8.7	222	22	Bacillus lichenifo
37	179	8.6	222	22	Bacillus lichenifo
38	179	8.6	222	22	Bacillus lichenifo
39	179	8.6	222	22	Bacillus lichenifo
40	179	8.6	222	22	Bacillus lichenifo
41	179	8.6	222	22	Bacillus lichenifo
42	179	8.6	222	22	Bacillus lichenifo
43	179	8.6	222	22	Bacillus lichenifo
44	179	8.6	222	22	Bacillus lichenifo
45	179	8.6	222	22	Bacillus lichenifo

ALIGNMENTS

RESULT 1

AA08657 1
ID AA08657 standard; Protein; 383 AA.

XX AC AA08657;

XX DT 09-AUG-1999 (first entry)

XX DE Human transmembrane domain containing protein from clone HP10493.

XX KW Transmembrane domain; human; nutrition; cytokine; cell differentiation;

XX KW immune stimulation; immune suppression; haematopoiesis; activin;

XX KW regulatory tissue growth; inhibin; chemostatic; chemokinetic;

XX KW haemostatic; thrombolytic; tumour inhibitor; anti-inflammatory;

XX KW gene therapy; screening.

XX OS Homo sapiens.

XX PN WO9927094-A2.

XX PD 03-JUN-1999.

XX PF 20-NOV-1998; 98WO-JP05238.

XX PR 25-NOV-1997; 97JP-0323129.

XX PA (PROT-) PROTEGENE INC.

XX PA (SAGA) SAGAMI CHEM RES CENT.

XX PI Kato S, Kimura T, Sekine S;

XX DR WPI: 1999-357835/30.

XX DR N-PSDB; AAX77690, AAX77691.

PT Novel proteins containing transmembrane domains, useful as
 PT anti-inflammatories, immune stimulators/suppressors and tissue
 XX growth compounds

XX Claim 1; Page 68-69; 89pp; English.

XX This invention describes novel human transmembrane containing proteins
 CC and their encoding nucleic acids. Although no specific use is given for
 CC the proteins, they may have a range of activities selected from
 CC nutritional uses, cytokine and cell differentiation, immune
 CC stimulation/suppression, haematopoiesis regulatory, tissue growth,
 CC activin/inhibin, chemostatic/chemokinetic, haemostatic/thrombolytic,
 CC receptor/ligand, tumour inhibitor, anti-inflammatory and other undefined
 CC activities. The cDNAs can be utilized as probes for gene diagnosis and
 CC as gene sources for gene therapy. The cDNAs can also be used for large
 CC scale expression of proteins. The transformed cells can be used for
 CC detection of the corresponding ligands and for screening of novel
 CC low-molecular pharmaceuticals.

XX Sequence 383 AA;

Query Match 100.0%; Score 2080; DB 20; Length 383;
 Best Local Similarity 100.0%; Pred. No. 1e-146;
 Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGIPGLLFFLLCAVGQVSPYSAPKPTWPAYRLPVVLPQSTLNLAKEPDKFAEAKLE 60
 DB 1 magipglilffllcavgvspysapkwptwpayrlpvlpqstlnlakpfdgaeakle 60
 QY 61 VSSCGPQCHKGTPLYEAKQYLSYETLYANGSRTEQVGIYILSSGGAQHRS 120
 DB 61 vsscgpqchkgtpltyeeakqylsyetlyangsrteqvgiylsssgdgahrsdgs 120
 QY 121 SGKRRRKQIYGVDSRFSIEFKDFLLNYPSTSVKLSGTGCTLVAEKHVLTAAHCIDHG 180
 DB 121 sgksrrrkqiygydsrfsifgkdfllnypstsvklsstgctgtlvaeakhvltaaahchidg 180
 QY 181 KTYVKGTKLRVGLPKPKFKDGGANDSTSAMPEQMKFQWIRVTRHPKVGKIGNAND 240
 DB 181 ktyvkgtklrvglpkpkfkdggrgandstsampeqmkfqwirvtrhvpkgwkgignand 240
 QY 241 IGMDDYVALLLELKKPKRKFVKIGVSPAKQLPGGRTHFSGYDNDRPNLVYFCVDE 300
 DB 241 igmddyvallyellkpkhrkfkmgkigvspakqlpggrthfsgydnrdpnlvyrfcvde 300
 QY 301 TYDLLYQCCDAQPCAGSGYVVRMWRKQKWERKIIGIFSGHGWDMNGSPQDFNVAVR 360
 DB 301 tydillyqccdaqpgasgvyvrmwkrqgqkwerkiigifsgghqwdmngspqdfnvavr 360
 QY 361 ITPLKYAQICYWIKGNVLDREG 383
 DB 361 itplkyaqicywikgnvldreg 383

RESULT 2

AAV08660
 ID AAV08660 standard; Protein; 383 AA.

XX AC

XX AC

XX DT 09-AUG-1999 (first entry)

XX DE W09927094 Seq ID 12.

XX KW

XX Transmembrane domain; human; nutrition; cytokine; cell differentiation;
 KW immune stimulation; immune suppression; haematopoiesis; activin;
 KW regulatory tissue growth; inhibin; chemostatic; chemokinetic;
 KW haemostatic; thrombolytic; tumour inhibitor; anti-inflammatory;
 KW gene therapy; screening.

XX OS

XX HOMO sapiens.

XX

PN W09927094-A2.
 XX PD 03-JUN-1999.
 XX PF 20-NOV-1998; 98WO-JP05238.
 XX PR 25-NOV-1997; 97JP-0323129.
 XX (PROT-) PROTEGENE INC.
 PA (SAGA) SAGAMI CHEM RES CENT.
 XX PI Kato S, Kimura T, Sekine S;
 XX WPI; 1999-357835/30.

XX Novel proteins containing transmembrane domains, useful as
 PT anti-inflammatories, immune stimulators/suppressors and tissue
 PT growth compounds

XX PS Disclosure; Page 87-89; 89pp; English.

XX This invention describes novel human transmembrane containing proteins
 CC and their encoding nucleic acids. Although no specific use is given for
 CC the proteins, they may have a range of activities selected from
 CC nutritional uses, cytokine and cell differentiation, immune
 CC stimulation/suppression, haematopoiesis regulatory, tissue growth,
 CC activin/inhibin, chemostatic/chemokinetic, haemostatic/thrombolytic,
 CC receptor/ligand, tumour inhibitor, anti-inflammatory and other undefined
 CC activities. The cDNAs can be utilized as probes for gene diagnosis and
 CC as gene sources for gene therapy. The cDNAs can also be used for large
 CC scale expression of proteins. The transformed cells can be used for
 CC detection of the corresponding ligands and for screening of novel
 CC low-molecular pharmaceuticals.

XX Sequence 383 AA;

Query Match 100.0%; Score 2080; DB 20; Length 383;
 Best Local Similarity 100.0%; Pred. No. 1e-146;
 Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGIPGLLFFLLCAVGQVSPYSAPKPTWPAYRLPVVLPQSTLNLAKEPDKFAEAKLE 60
 DB 1 magipglilffllcavgvspysapkwptwpayrlpvlpqstlnlakpfdgaeakle 60
 QY 61 VSSCGPQCHKGTPLYEAKQYLSYETLYANGSRTEQVGIYILSSGGAQHRS 120
 DB 61 vsscgpqchkgtpltyeeakqylsyetlyangsrteqvgiylsssgdgahrsdgs 120
 QY 121 SGKRRRKQIYGVDSRFSIEFKDFLLNYPSTSVKLSGTGCTLVAEKHVLTAAHCIDHG 180
 DB 121 sgksrrrkqiygydsrfsifgkdfllnypstsvklsstgctgtlvaeakhvltaaahchidg 180
 QY 181 KTYVKGTKLRVGLPKPKFKDGGANDSTSAMPEQMKFQWIRVTRHPKVGKIGNAND 240
 DB 181 ktyvkgtklrvglpkpkfkdggrgandstsampeqmkfqwirvtrhvpkgwkgignand 240
 QY 241 IGMDDYVALLLELKKPKRKFVKIGVSPAKQLPGGRTHFSGYDNDRPNLVYFCVDE 300
 DB 241 igmddyvallyellkpkhrkfkmgkigvspakqlpggrthfsgydnrdpnlvyrfcvde 300
 QY 301 TYDLLYQCCDAQPCAGSGYVVRMWRKQKWERKIIGIFSGHGWDMNGSPQDFNVAVR 360
 DB 301 tydillyqccdaqpgasgvyvrmwkrqgqkwerkiigifsgghqwdmngspqdfnvavr 360
 QY 361 ITPLKYAQICYWIKGNVLDREG 383
 DB 361 itplkyaqicywikgnvldreg 383

RESULT 3

AAV13390

ID AAV13390 standard; Protein; 383 AA.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:25:50 ; Search time 84.84 Seconds
(without alignments)
660.329 Million cell updates/sec

Title: US-09-072-384-18

Perfect score: 2080.

Sequence: 1 MAGIPGLLEFLFLCAVQ.....LKAYICYWIKGNLYDCREG 383

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 14627329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL17:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2080	100.0	383	4	Q95084
2	1884.5	90.6	382	11	Q9D6X6
3	1042	50.1	413	4	Q9BQP6
4	133.5	6.4	799	11	Q9DB10
5	129.5	6.2	303	2	Q9EXR9
6	129	6.2	469	6	Q9GMD9
7	127.5	6.1	1322	5	Q9NAT0
8	126.5	6.1	678	11	Q9JJS8
9	124.5	6.0	1322	5	Q9NJ55
10	124	6.0	266	6	Q46644
11	123.5	5.9	339	11	Q9QX91
12	123.5	5.9	366	11	Q9QX85
13	123.5	5.9	541	11	Q9QX90
14	123.5	5.9	623	11	Q9JJP3
15	123.5	5.9	643	11	Q9QX84
16	122	5.9	259	5	Q9XY61
17	122	5.9	449	5	Q9VD08
18	120.5	5.8	482	11	Q63207
19	118	5.7	1376	5	Q9VQR8

20	117	5.6	258	5	Q9W508
21	117	5.6	522	5	O01771
22	115.5	5.6	266	11	Q9D936
23	115.5	5.6	269	4	Q14243
24	115	5.5	258	4	Q9UN11
25	113.5	5.5	418	5	Q9VA87
26	113.5	5.5	685	11	Q9Z338
27	113.5	5.5	737	13	Q90422
28	113	5.4	573	5	Q9V516
29	112.5	5.4	603	5	Q9VDV1
30	112.5	5.4	686	4	Q9BZH0
31	112	5.4	249	13	Q9W7Q1
32	110.5	5.3	279	5	Q9G991
33	110.5	5.3	481	11	O54740
34	110	5.3	1449	5	Q10922
35	109.5	5.3	405	2	Q69973
36	108	5.2	274	5	Q17086
37	108	5.2	750	13	Q9W633
38	108	5.2	1047	5	Q24019
39	107.5	5.2	236	11	Q9Z1H1
40	107.5	5.2	427	10	Q9AYR4
41	107.5	5.2	481	11	O88947
42	107.5	5.2	481	11	Q99L32
43	107.5	5.2	611	5	Q9VZS8
44	106	5.1	460	5	Q9VFZ6
45	106	5.1	762	13	Q9YIC6

ALIGNMENTS

RESULT 1

ID O95084 PRELIMINARY; PRT; 383 AA.

AC O95084;

DT 01-MAY-1999 (TREMBlrel. 10, Created)

DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE SERINE PROTEASE (HYPOTHETICAL 43.0 KDA PROTEIN) (PROTEASE, SERINE, DE 23).

GN ZSG13 OR DKFZP586B0719.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=UMBILICAL VEIN;

RA Li X., Tedder T.F.;

RT "A novel serine protease from human umbilical vein endothelial cells.";

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Sheppard P., Blumberg H., Jelinek L., Foster D., O'Hara P.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE=UTERUS;

RA Wambutt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RC TISSUE=CERVIX CARCINOMA;

RA Strausberg R.;

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.

CC -!- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).

DR EMBL; AF015287; AAD01553.1; -

DR EMBL; AF193611; AAF07186.1; -

DR EMBL; AF136914; CAB66848.1; -

DR EMBL; BC001278; AAH01278.1; -

DR MEROPS; S01.309; -

```

DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYPSIN_1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 383 AA; 43001 MW; 46EB6C11ABFD58F CRC64;

Query Match 100.0%; Score 2080; DB 4; Length 383;
Best Local Similarity 100.0%; Pred. No. 8.2e-184;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGIPGLFLFLFLLCAVGQVSPYAPWKTWPAYRLPVVLPVLPSTLNLAKEPFGAEAKLE 60
Db 1 MAGIPGLFLFLFLLCAVGQVSPYAPWKTWPAYRLPVVLPVLPSTLNLAKEPFGAEAKLE 60

QY 61 VSSSCGQCHKGTPPLYEAKQVLSYETLYANGSRTEQVGIYILSSSGDGAQHRDGS 120
Db 61 VSSSCGQCHKGTPPLYEAKQVLSYETLYANGSRTEQVGIYILSSSGDGAQHRDGS 120

QY 121 SGKSRKRLQIYGYDSRFSIFGKDFLLNYPSTSVKLTGCTGLVAEKHVLTAACHIDG 180
Db 121 SGKSRKRLQIYGYDSRFSIFGKDFLLNYPSTSVKLTGCTGLVAEKHVLTAACHIDG 180

QY 181 KTVYKGTQKLRVGLFKPKFKDGGANDSTSAMPEQMKFQWIRKRVHPKGIKGNAND 240
Db 181 KTVYKGTQKLRVGLFKPKFKDGGANDSTSAMPEQMKFQWIRKRVHPKGIKGNAND 240

QY 241 IGMDDYALLELKKPKHKKPMKIGVSPAPKQLPGRHIFSGYDNDRPNLYVRFCDVKDE 300
Db 241 IGMDDYALLELKKPKHKKPMKIGVSPAPKQLPGRHIFSGYDNDRPNLYVRFCDVKDE 300

QY 301 TYDLYQQCDAQPGASGSGVYVWRMKRQOKWERRKIIGIFSGHQQWDMNGSPQDFNVAVR 360
Db 301 TYDLYQQCDAQPGASGSGVYVWRMKRQOKWERRKIIGIFSGHQQWDMNGSPQDFNVAVR 360

RESULT 2
Q9B6X6 PRELIMINARY; PRT; 382 AA.
ID Q9B6X6
AC Q9B6X6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 2310046G15RIK PROTEIN.
GN 2310046G15RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TONGUE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

```

```

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC -!- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (SI).
DR EMBL: AK009847; BAB26541.1; -.
DR MGD; MGI:1923703; 2310046G15RIK.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYPSIN_1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 382 AA; 43053 MW; 69C9A7080E5B2306 CRC64;

Query Match 90.6%; Score 1884.5; DB 11; Length 382;
Best Local Similarity 90.6%; Pred. No. 9.2e-166;
Matches 347; Conservative 15; Mismatches 20; Indels 1; Gaps 1;

QY 1 MAGIPGLFLFLFLLCAVGQVSPYAPWKTWPAYRLPVVLPVLPSTLNLAKEPFGAEAKLE 60
Db 1 MAGIPG-LFILLVLLCVFMQVSPYTPVWKTWPAYRLPVVLPVLPSTLNLAKEPFGAEAKLE 59

QY 61 VSSSCGQCHKGTPPLYEAKQVLSYETLYANGSRTEQVGIYILSSSGDGAQHRDGS 120
Db 60 VSSSCGQCHKGTPPLYEAKQVLSYETLYANGSRTEQVGIYILSSSGDGAQHRDSEA 119

QY 121 SGKSRKRLQIYGYDSRFSIFGKDFLLNYPSTSVKLTGCTGLVAEKHVLTAACHIDG 180
Db 120 TGRSRKRLQIYGYDGRFSIFGKDFLLNYPSTSVKLTGCTGLVAEKHVLTAACHIDG 179

QY 181 KTVYKGTQKLRVGLFKPKFKDGGANDSTSAMPEQMKFQWIRKRVHPKGIKGNAND 240
Db 180 KTVYKGTQKLRVGLFKPKFKDGGANDSSSSAMPDKMKFQWIRKRVHPKGIKGNAND 239

QY 241 IGMDDYALLELKKPKHKKPMKIGVSPAPKQLPGRHIFSGYDNDRPNLYVRFCDVKDE 300
Db 240 IGMDDYALLELKKPKHKKPMKIGVSPAPKQLPGRHIFSGYDNDRPNLYVRFCDVKDE 299

QY 301 TYDLYQQCDAQPGASGSGVYVWRMKRQOKWERRKIIGIFSGHQQWDMNGSPQDFNVAVR 360
Db 300 TYDLYQQCDAQPGASGSGVYVWRMKRQOKWERRKIIGIFSGHQQWDMNGSPQDFNVAVR 359

QY 361 ITPLKYAICYWIKNYLDREG 383
Db 360 ITPLKYAICYWIKNYLDREG 382

RESULT 3
Q9BQP6 PRELIMINARY; PRT; 413 AA.
ID Q9BQP6
AC Q9BQP6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DJ223E3.1 (PUTATIVE SECRETED PROTEIN Z5IG13).
GN DJ223E3.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC -
RX -
RA Dunn M.;
RA Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL: AL121939; CAC35071.1; -.
DR EMBL; AL121939; CAC35071.1; -.
SQ SEQUENCE 413 AA; 47098 MW; 818D9C951BD2D6C1 CRC64;

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:24:07 ; Search time 32.13 Seconds
(without alignments)
447.327 Million cell updates/sec

Title: US-09-072-384-15

Perfect score: 2131

Sequence: 1 MAGIPGLFLFLFLCAVQ.....IKGYLDRCRGDTVPFPGSN 392

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	194	9.1	316	1	GSEP_BACLI
2	135.5	6.4	313	1	MPR_BACSU
3	124	5.8	286	1	EL1_BOVIN
4	121.5	5.7	490	1	FA10_RABIT
5	119	5.6	269	1	EL2_PIG
6	115.5	5.4	269	1	EL2A_HUMAN
7	115.5	5.4	271	1	EL2_MOUSE
8	115	5.4	376	1	FA10_TROCA
9	114.5	5.4	238	1	TRY5_AEDAE
10	113	5.3	266	1	EL1_PIG
11	113	5.3	266	1	EL1_RAT
12	112.5	5.3	259	1	TRYP_STRGR
13	112.5	5.3	273	1	YGD_ECOLI
14	111.5	5.2	488	1	FA10_HUMAN
15	111.5	5.2	875	1	NETR_HUMAN
16	111	5.2	253	1	CFAD_HUMAN
17	111	5.2	761	1	NETR_MOUSE
18	110	5.2	269	1	EL2B_HUMAN
19	110	5.2	492	1	FA10_BOVIN
20	108	5.1	583	1	CFAL_HUMAN
21	108	5.1	786	1	STUB_DROME
22	107.5	5.0	686	1	MAS2_HUMAN
23	107	5.0	269	1	EL2_BOVIN
24	107	5.0	603	1	CFAL_MOUSE
25	105.5	5.0	604	1	CFAL_RAT
26	105	4.9	274	1	TRY5_ANOGA
27	104.5	4.9	268	1	TRYP_STRGA
28	103.5	4.9	271	1	EL2_RAT
29	103.5	4.9	416	1	FA9_BOVIN
30	103	4.8	855	1	SL14_HUMAN
31	102	4.8	855	1	SL14_MOUSE
32	101.5	4.8	281	1	TRY2_DROER
33	98.5	4.6	260	1	COGS_HYPLI

34 98.5 4.6 267 1 TRY7_ANOGA
35 97.5 4.6 452 1 FA9_CANFA
36 97 4.6 1019 1 LFC_TACTR
37 96.5 4.5 252 1 TRY1_DROME
38 96.5 4.5 259 1 DEF3_DERE
39 96.5 4.5 275 1 FA9_RABIT
40 96.5 4.5 454 1 TMS3_HUMAN
41 96.5 4.5 2145 1 CYAA_PODAN
42 96 4.5 274 1 TRY1_ANOGA
43 95.5 4.5 430 1 SNAK_DROME
44 95 4.5 1019 1 LFC_CARRO
45 94.5 4.4 261 1 DER3_DERPT

ALIGNMENTS

RESULT 1

GSEP_BACLI
ID GSEP_BACLI STANDARD: PRT: 316 AA.
AC P80057;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GLUTAMYL ENDOPEPTIDASE PRECURSOR (EC 3.4.21.19) (GLUTAMATE SPECIFIC
DE ENDOPEPTIDASE) (GSE).
GN BLASE.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 14580;
RX MEDLINE=93054737; PubMed=1429718;
RA Kakudo S., Kikuchi N., Kitadokoro K., Fujiwara T., Nakamura E.,
RA Okamoto H., Shin M., Tamaki M., Teraoka H., Tsuzuki H., Yoshida N.;
RT "Purification, characterization, cloning, and expression of a
RT glutamic acid-specific protease from Bacillus licheniformis ATCC
RT 14580.";
RL J. Biol. Chem. 267:23782-23788(1992).
RN [2]
RP SEQUENCE OF 95-316.
RX MEDLINE=92155199; PubMed=1346764;
RA Svendsen I., Bredam K.;
RT "Isolation and amino acid sequence of a glutamic acid specific
RT endopeptidase from Bacillus licheniformis.";
RL Eur. J. Biochem. 204:165-171(1992).
CC -!- FUNCTION: SPECIFIC FOR HYDROLYSIS OF PEPTIDES BONDS ON THE
CC CARBOXYL SIDE OF ACIDIC AMINO ACID RESIDUES, WITH A STRONG
CC PREFERENCE FOR GLU.
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ASP-I-XAA, GLU-I-XAA.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE
CC V8 FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: D10060; BAA00949.1; -
PIR: S23078; S23078.
PIR: A45134; A45134.
MEROPS: S01.271; -
InterPro: IPR00126; Ser_proteas_v8.
InterPro: IPR001254; Trypsin.
Pfam: PF00089; trypsin; 1
PRINTS: PR00839; V8PROTEASE.
SMART: SM00020; Tryp_Spc; 1.

DR PROSITE; PS00672; V8_HIS; 1.
KW Hydrolyase; Serine protease; Signal.
FT SIGNAL 1 ?
FT PROPEP 1 ?
FT CHAIN 95 316 GLUTAMYL ENDOPEPTIDASE.
FT ACT_SITE 141 141 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 261 261 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 126 142
FT DISULFID 275 279
SQ SEQUENCE 316 AA; 96D7552CB7089B09 CRC64;

Query Match 9.1%; Score 194; DB 1; Length 316;
Best Local Similarity 24.0%; Pred. No. 5e-09;
Matches 79; Conservative 43; Mismatches 143; Indels 64; Gaps 15;

QY 63 SSCGPGCHKTPL---PTYEAKQYLSYEITYANGSRTEQTQVGIYILSSSGDGAHQHDSG 119
DB 28 AQAAPSPH--TPVSSDPSY-KAETSVDYDP-----NIKSDQGLYSKAFGTGKVNTEKE 79

QY 120 SSGSRKRQIYGDYSRFSIFGKDFLEN-----YPFSTSVKLSST---GCTGTILVAEKHV 170
DB 80 KAERKSPAKAPY---SIKSYIGSDRTRVNTATPYEAIHVHSSISGSGCTGWMIGPKTV 136

QY 171 LTAACHIDGKT-VYKQTKLRVGLPKPKDGRGANDSTSAMPEQMFKQWIRKRVTHV 229
DB 137 ATAGHCITYDTSSGFACTATVSPG-----RNGTST-----YPGSVKSTRYFI 178

QY 230 PKGWIKGNANDIGMDYDYLLELKKPKHFKEMKIGVSPPAKQLPGGRIHPSGYDNDPRGN 289
DB 179 PSGRNSNTN-----YDGAIELSEPIGNTVGVGYSTTSSLVGTVTITSGYPCDRTAG 233

QY 290 LVYRFCD--VKDETIDLYOCCDAQPCAGSGGVVVRWKKRQOKWERKII---GIFSGH 343
DB 234 TQWHSGFIAISEYIKLOYAM-DTYGQSGSPVFEQSSSRNCSGCSLAVHTNGVYGG- 291

QY 344 QWYDMNGSPQDNVAVRITPLKVAICYW 372
DB 292 -----SSYNGRTRITKEVFDNLNW 311

RESULT 2
MPR_BACSU STANDARD; PRT; 313 AA.
AC P39790;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EXTRACELLULAR METALLOPROTEASE PRECURSOR (EC 3.4.21.-).
GN MPR.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=GP241;
RX MEDLINE=90130256; PubMed=2105291;
RA Sloma A., Rudolph C.F., Rufo G.A. Jr., Sullivan B.J., Theriault K.A., Ally D., Pero J.;
RT "Gene encoding a novel extracellular metalloprotease in Bacillus subtilis.";
RL J. Bacteriol. 172:1024-1029(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Haga K., Liu H., Yasumoto K., Takahashi H., Yoshikawa H.;
RT "Sequence analysis of the 70kb region between 17 and 23 degree of the Bacillus subtilis chromosome.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-68 FROM N.A.

RX MEDLINE=89108019; PubMed=3145906;
RA Smith H., de Jong A., Bron S., Venema G.;
RT "Characterization of signal-sequence-coding regions selected from the Bacillus subtilis chromosome.";
RL Gene 70:351-361(1988).
RN [4]
RP SEQUENCE OF 1-10 FROM N.A.
RC STRAIN=168;
RX MEDLINE=961118702; PubMed=7496533;
RA Saxild H.H., Jacobsen J.H., Nygaard P.;
RT "Functional analysis of the Bacillus subtilis purT gene encoding formate-dependent glycineamide ribonucleotide transferase.";
RL Microbiology 141:2211-2218(1995).
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE V8 FAMILY.
CC -1- CAUTION: CALLED "METALLOPROTEASE", BUT CLEARLY BELONGS TO THE S2 FAMILY OF SERINE PROTEASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; L10505; AAA22604.1; -
CC EMBL; AB006424; BAA33121.1; -
CC EMBL; M22916; AAA22832.1; -
CC EMBL; Z99105; CAB12018.1; -
CC PIR; A35122; A35122.
CC HSP; P00756; 1SGF.
CC MEROPS; S01.272; -
CC Subtilist; BGI0690; mpr.
CC Interpro; IPR000126; Ser_proteas_V8.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; P00839; V8PROTEASE.
CC SMART; SM00020; Tryp_SPC; 1.
CC PROSITE; PS00672; V8_HIS; 1.
CC PROSITE; PS00673; V8_SER; 1.
KW Hydrolyase; Serine protease; Signal; Zymogen; Complete proteome.
FT SIGNAL 1 34
FT PROPEP 35 93 EXTRACELLULAR METALLOPROTEASE.
FT CHAIN 94 313 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 146 146 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 267 267 BY SIMILARITY.
FT DISULFID 131 147
FT CONFLICT 61 68. QVSAPYEG -> PLESTAQA (IN REF. 3).
SQ SEQUENCE 313 AA; 33842 MW; D41788B8D652AE94 CRC64;

Query Match 6.4%; Score 135.5; DB 1; Length 313;
Best Local Similarity 22.2%; Pred. No. 0.00038;
Matches 77; Conservative 38; Mismatches 121; Indels 111; Gaps 17;

QY 72 GTPLPYEEAKQYLSYEITYANGSRTEQTQVGIYILSSSGD--GAQHRDSSGSKSRKRQ 129
DB 29 GVPAKAENQTSVNTGKKEADTKNOT-----SKADQVSAPYEKTKTSK----- 75

QY 130 IYGYDSRF-----SIFGKD-----FLNYPFSTSVKLSST-----GCTGTL 164
DB 76 LYGGQTELEKNIQTLPSSIICTDERTRISSTTFPYRATVQLSIKYPNTSSYVGCTGFL 135
QY 165 VAEKHVLTAACHIH-----DGKTYVKGTKQLRVGLPKPKFDGGRGAND 208
DB 136 VNPNTVVTAGCHVSDHGWASTITAAPGRNGSSYPYGT----- 175
QY 209 STSAMPEQMFKQWIRKRVHVPKWKGN---ANDIGMDYDYLLELKKPKHFKMKIGV 265
DB 176 -SGTMEFYSVK-GWTESKDTNYDGAIKLNSPGNTVGVW-YGYRTNSSSP-----VGL 225

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:17:55 ; Search time 48.03 seconds
(without alignments)
621.704 Million cell updates/sec

Title: us-09-072-384-15

Perfect score: 2131

Sequence: 1 MAGIPGLFLFLFLCAVGO.....IKGNLYDCREGDTVFPPGSN 392

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5.

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_68.*

1: p1r1.*

2: p1r2.*

3: p1r3.*

4: p1r4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	194	9.1	316	2 A45134	endopeptidase (EC
2	135.5	6.4	313	2 A35122	metalloproteinase
3	120.5	5.7	482	1 EXRT	coagulation factor
4	119	5.6	269	2 A26823	pancreatic elastase
5	117	5.5	522	2 T29767	hypothetical prote
6	115.5	5.4	269	2 B26823	pancreatic elastase
7	115.5	5.4	271	2 A25528	pancreatic elastase
8	115	5.4	258	2 S70439	pancreatic elastase
9	115	5.4	267	4 A56615	probable pancreati
10	114.5	5.4	238	1 TRWVSY	trypsin-like prote
11	114	5.3	246	1 DBHU	complement factor
12	113	5.3	266	1 ELRT1	pancreatic elastase
13	113	5.3	266	1 ELPG	pancreatic elastase
14	112.5	5.3	259	1 TRSMG	trypsin (EC 3.4.21
15	112.5	5.3	273	2 E85765	hypothetical prote
16	112.5	5.3	273	2 H64915	hypothetical prote
17	111.5	5.2	488	1 EXHU	coagulation factor
18	111	5.2	761	2 JC5759	brain-specific ser
19	110	5.2	269	2 C26823	pancreatic elastase
20	110	5.2	492	1 EXBO	coagulation factor
21	110	5.2	1582	2 T15308	hypothetical prote
22	109.5	5.1	405	2 T35117	probable secreted
23	108	5.1	583	2 A29154	complement factor
24	108	5.1	786	1 A7547	serine proteinase
25	108	5.1	1047	2 A55617	masquerade precurs
26	107.5	5.0	236	2 A28566	T-cell suppressor
27	107.5	5.0	686	1 A59271	Ra-reactive factor
28	105	4.9	274	2 S40004	trypsin-related pr
29	103.5	4.9	271	1 ELRT2	pancreatic elastase

30	103.5	4.9	416	1 KFEBO	coagulation factor
31	102.5	4.8	1238	2 T34929	hypothetical prote
32	100	4.7	272	2 JC4170	trypsin-like prote
33	98.5	4.6	267	2 S40006	trypsin (EC 3.4.21
34	98	4.6	548	2 D82175	probable trypsin V
35	98	4.6	624	2 T02289	probable polygalac
36	97.5	4.6	409	2 T35118	coagulation factor
37	97.5	4.6	452	1 A30351	complement factor
38	97.5	4.6	747	2 I51579	coagulation factor
39	97	4.6	1019	2 A38738	allergen Der f III
40	96.5	4.5	259	2 S68424	factor IX - rabbit
41	96.5	4.5	275	2 I46712	adenylate cyclase
42	96.5	4.5	2145	2 JC4747	serine proteinase
43	95.5	4.5	430	1 A24702	hypothetical prote
44	95	4.5	782	2 T32155	pancreatic elastase
45	94	4.4	268	2 S68825	pancreatic elastase

ALIGNMENTS

RESULT 1

A45134

endopeptidase (EC 3.4.-.-), glutamate-specific - Bacillus licheniformis

C:Species: Bacillus licheniformis

C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999

C:Accession: A45134; S23078

R:Kakudo, S.; Kikuchi, N.; Kitadokoro, K.; Fujiwara, T.; Nakamura, E.; Okamoto, H.; S

J. Biol. Chem. 267, 23782-23786 (1992)

A:Title: Purification, characterization, cloning, and expression of a glutamic acid-s

A:Reference number: A45134; MUID:93054737

A:Accession: A45134

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-316 <KAK>

A:CROSS-references: GB:D10060; NID:g216263; PIDN:BAA00949.1; PID:d1001415; PID:g21626

A:Experimental source: ATCC 14580

A:Note: sequence extracted from NCBI backbone (NCBIN:118784, NCBIP:118785)

R:Svendsen, I.; Brøddam, K.

Eur. J. Biochem. 204, 165-171, 1992

A:Title: Isolation and amino acid sequence of a glutamic acid specific endopeptidase

A:Reference number: S23078; MUID:92155199

A:Accession: S23078

A:Status: preliminary

A:Molecule type: protein

A:Residues: 95-316 <SVE>

C:Keywords: hydrolase

Query Match 9.1%; Score 194; DB 2; Length 316;
Best Local Similarity 24.0%; Pred. No. 1.4e-08;
Matches 79; Conservative 43; Mismatches 143; Indels 64; Gaps 15;

QY 63 SCGQPCQCHGTPL---PYEEAKQVLSYETLYANGSRPTQVGIYILSSGGCAHQHDSG 119

DB 28 AQAAPSPH--TPVSSDPSY-KAETSVTDP-----NIKSDQYGLSKAFPTGKVNETKE 79

QY 120 SSGKRRKRIQYDYSRFSIFCKDFLN-----YFSTSVKLST---GCTGLVAEKHV 170

DB 80 KAEKSPAKAPY---SIKSVIGSDDRTRVNTATPYRAIVHISSIGSCTGWMGPKTV 136

QY 171-LPAHETHDCKT-YVKGQTKLVGFLKPKFKDGGGRANDSTSAMPEQMFKFOWIRVKRTHV 229

DB 137 AFAGHCIVDTSSGSPAGTATVSPG-----RNGTS-----YPYGVSKSTRVFI 178

QY 230 PGKWTGNANDIGMDYDVALLELKKPKFKMKGIVSPPAKOLPGRIHFSGYDNDPRGN 289

DB 179 PSGMRSGNTN-----YDGAIELSEPIGNTVGYFGYSYTTSSLVGTVTITISGPGDKTAG 233

QY 290 LVYRCD--YKQETVLLVQCDAPGASGSGVYVRMKRKROOKWERKII-----GIFSCH 343

DB 234 TQWQHSGPIASETKQLQAM-DTYGGSGSPVFPQSSSRNTCSGPCSLAVHTNGVYGG- 291

Qy 344 QWDMNGSPQDENVAVRITPLKYAQICYW 372
 Db 292 -----SSYNRGTRITKEVFDNLTNW 311

RESULT 2

A35122

metalloprotease (EC 3.4.21.1) mpr precursor, extracellular - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 27-Jul-1990 #sequence_revision 27-Jul-1990 #text_change 15-Oct-1999

C:Accession: A35122; I40010; A69660

R:Sloma, A.; Rudolph, C.F.; Rufo Jr., G.A.; Sullivan, B.J.; Theriault, K.A.; Ally, D.; F.

J. Bacteriol. 172, 1024-1029, 1990

A:Title: Gene encoding a novel extracellular metalloprotease in Bacillus subtilis.

A:Reference number: A35122; MUID:90130256

A:Accession: A35122

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-313 <SILO>

A:Cross-references: GB:110505; NID:gl43209; PIDN:AAA22604.1; PID:gl43210; GB:M29036

R:Smith, H.; de Jong, A.; Bron, S.; Venema, G.

Gene 70, 351-361, 1988

A:Title: Characterization of signal-sequence-coding regions selected from the Bacillus s

A:Reference number: I39994; MUID:89108019

A:Accession: I40010

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-60,65, 'L',67, 'S',69, 'AQA' <RES>

A:Cross-references: GB:M22916; NID:gl43701; PIDN:AAA22832.1; PID:gl43702

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter

C: Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch

A: Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gallen

teck, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidos, A.; Lardinios,

A:Authors: Leuber, K.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauel

Y, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tonponi, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MUID:98044033

A:Accession: A69580

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

```

Query Match          6.4%; Score 135.5; DB 2; Length 313;
Best Local Similarity 22.2%; Pred.-No. 0.001;
Matches 77; Conservative 38; Mismatches 121; Indels 111; Gaps 17;

QY 72 GTPLPTTVEEAKQYLSEYLLYANGSRRTQGVILLSSGD--GAQHRDSCSSGSKRRKQ 129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29 GVPKAAENPOTSNTGKEADATKNOT-----SKAQVSAPYEGTGTKTSK----- 75

QY 130 IYGYDSRF-----SIFGKD-----FLNYPFSTSVKLST-----GCTGTL 164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76 LYGGOTELEKNIQTLQPSIIIGTDERTRISSTTSFPYRATVQLSIKYPNTSSYVCGTGFL 135

QY 165 VAEKHVLTAACHT-----DGYTYVKGTOKL RVGFLKPKFKDGGRGAND 208
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 166 VNPNTVTAGHCYVSQDHGWASTITAAPGRNGSSYPYGT----- 175

QY 209 STSAMPEQMKFQWIRVKRTHVHPKGIKN---ANDIGMDYDYALLKLPKHKRKMKGIV 265

```

Db 176 -SGTMFYSVK-GWTFESKDTNYDYGAIKLNGSPGNTVGW-YGYRTNTSSP-----VGL 225
QY 266 SPPAKQLPGGRIHSPGNDNRPGNLVRYFCDVKDTEYDLLLYQQCDAQPGASGSGVYVRMW 325
Db 226 SSSVTGFPCKDTFGTMSDKPIR-----SAETVKLVY-TTDTYGCQSGSPVY---- 272
QY 326 KRQOQKWERKLIIFSGHQWDMNGSPQDENVAVRTPLPKYAICYW 372
Db 273 -RNYSDTGQTAAIAHT-----NGG-SSYNLGRVTVNDVFNNIQYW 310
RESULT 3
EXRT
coagulation factor Xa (EC 3.4.21.6) precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Jan-1995 #sequence_revision 07-Feb-1997 #text_change 08-Dec-2000
C:Accession: S49075; JC4670; PS0191; PS0190; I62745
R:Stanton, C.; Ross, P.; Hutson, S.; Wallin, R.
Thromb Res 80, 63-73, 1995
A:Title: Evidence for competition between vitamin K-dependent clotting factors
A:Reference number: A58498; MUID:96093366
A:Accession: S49075
A:Molecule type: mRNA
A:Residues: 1-482 <STA>
A:CROSS-references: EMBL:X79807; NID:g506600; PIDN:CAA56202.1; PID:g506601
A:Note: submitted to the EMBL Data Library June 1994
A:Note: neither the complete nucleic acid sequence nor the complete translation are
R:Stanton, C.; Ross, R.P.; Hutson, S.; Wallin, R.
Gene 169, 269-273, 1996
A:Title: Processing and expression of rat and human clotting factor-X-encoding cDNAs
A:Reference number: JC4670; MUID:96194815
A:Accession: JC4670
A:Molecule type: mRNA
A:Residues: 1-482 <STA>
A:CROSS-references: EMBL:X79807; NID:g506600; PIDN:CAA56202.1; PID:g506601
A:Experimental source: Cos-1 cell
R:Enyoji, K.; Miyazaki, K.; Kato, H.
J. Biochem. 109, 890-898, 1991
A:Title: Characterization of rat factors X and Xa: demonstration of factor Xa in rat
A:Reference number: PS0190; MUID:92041742
A:Accession: PS0191
A:Molecule type: protein
A:Residues: 41-58, 'X', 60-65 <ENJ1>
A:Accession: PS0190
A:Molecule type: protein
A:Residues: 183-186, 'X', 188-207 <ENJ2>
R:Murakawa, M.; Okamura, T.; Kamura, T.; Kuroiwa, M.; Harada, M.; Niho, Y.
Eur. J. Haematol. 52, 162-168, 1994
A:Title: Analysis of the partial nucleotide sequences and deduced primary structures
A:Reference number: I46196; MUID:94222160
A:Accession: I62745
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Residues: 295-383 'G', 385-455 <MUR>
A:CROSS-references: GB:D21215; NID:g415309; PIDN:BAA04756.1; PID:g455396
C:Function:
A:Description: catalyzes the proteolytic activation of prothrombin to thrombin in the
A:Pathway: blood coagulation
C:Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homol
C:Keywords: beta-hydroxyaspartic acid; blood coagulation; calcium binding; carboxyglu
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-40/Domain: propeptide #status predicted <PRO>
F:25-84/Domain: Gla domain homology <GLA>
F:41-179/Product: coagulation factor X light chain #status predicted <LCH>
F:90-121/Domain: EGF homology <EGI>
F:129-164/Domain: EGF homology <EC2>
F:183-482/Product: coagulation factor X heavy chain #status predicted <HCH>
F:183-231/Domain: activation peptide #status predicted <APT>
F:232-482/Product: coagulation factor Xa heavy chain #status predicted <ACT>
F:232-460/Domain: trypsin homology <TRY>
F:46, 47, 54, 56, 59, 60, 65, 66, 69, 72, 79/Modified site: gamma-carboxyglutamic acid (Glu) #s
F:57-62, 90-101, 95-110, 112-121, 129-149, 136-149, 151-164, 172-340, 238-243, 259-275, 388-402


```

QY 148 YPFSTSVK-LSYG-----CTGTIVAETHVLTAAHCIDHGKTY--VKGTOKLRVGFLPKPKF 199
      :: : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 42 WPQVQSLOVLSSRWNRNCGSLVANWLVLTAAHCLLNRYQTYRVLGHAHL----- 92
      :: : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 200 KDGGRGANDSTSAMPEQMKFOWIRVKTHVPKGWIKGNANDIGMDYDYALLELKKP-HKR 258
      || : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 93 --SNPGAGSAA-----VOVSKLVVHQHW---NSQNVNGCYDIALIKLASPVTL 136
      || : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 259 KFMKIGVSPPAKQ-LPGGRI-HFSGY-----DNDRPGNLV-----YRFC----- 295
      || : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 137 KNIOACLPAPAGTILPRNYCVYVTGWGLLOTNGNSPOTLRQGRLVVVDYATCSSASWWS 196
      || : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 296 DVKDETY----DLIYOQCDQAOPGAGSGSVYRMWKROOKWERKLIIFGSHQWDHNGS 351
      || : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 197 SVKSSMCWAGGDGVTTSCN---GDSGGPLNCRASNGO---WQ---VHGIVSFGSLGCN-Y 247
      || : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 352 PQDENNAVVRITPLKYAQCWIKNYILD 379
      || : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 248 PRKPSVFTRVS-----NYID 262
      || : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 8
S70439
pancreatic elastase I (allele HELI-16) probable splice form I - human
C:Species: Homo sapiens (man)
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 16-Dec-1998
C:Accession: S70439
R:kawashima, I.; Tani, T.; Mita-Honjo, K.; Shimoda-Takano, K.; Ohmine, T.; Furukawa,
DNA Seq. 2, 303-312, 1992
A:Title: Genomic organization of the human homologue of the rat pancreatic elastase I
A:Reference number: A56615; MUID:92338395
A:Accession: S70439
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-258 <RAW>
C:Superfamily: trypsin; trypsin homology
F:19-251/domain: trypsin homology <TRY>

Query Match 5.4%; Score 115; DB 2; Length 258;
Best Local Similarity 26.9%; Pred No. 0.04;
Matches 35; Conservative 23; Mismatches 38; Indels 34; Gaps 6;

QY 147 NYPFSTSVKLSTG-----CTGTIVAETHVLTAAHCIDHGKTY--VKGTOKLRVGFLPKPK 198
      :: : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 29 SWPSQISLQYSRGGSWHTCGTILRNQWMTAAHCVDYQKTFERWAGDHNL----- 80
      :: : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 199 FKDGGRGANDSTSAMPEQMKFOWIRVKTHVPKGWIKGNANDIGMDYDYALLEL-KKPHK 257
      || : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 81 -----SONDGE-----QYVSQKRVVHPYW---NSDNVAAGYDIALLRLAQSVTL 123
      || : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 258 RKFMKIGVSP 267
      :: : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 124 NSYVOLGVLP 133
      :: : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 9
A56615
probable pancreatic elastase (EC 3.4.21.36) pseudogene - human
N:Alternate names: pancreatic elastase I homolog; pancreatic elastase I allele HELI-1
C:Species: Homo sapiens (man)
C:Date: 11-Aug-1995 #sequence_revision 17-Aug-1995 #text_change 14-Aug-1998
C:Accession: A56615; S70440
R:kawashima, I.; Tani, T.; Mita-Honjo, K.; Shimoda-Takano, K.; Ohmine, T.; Furukawa,
DNA Seq. 2, 303-312, 1992
A:Title: Genomic organization of the human homologue of the rat pancreatic elastase I
A:Reference number: A56615; MUID:92338395
A:Accession: A56615
A:Molecule type: DNA
A:Residues: 1-267 <RAW>
A:Cross-references: EMBL:X62259; NID:g31246; EMBL:X62258; GB:S40923; NID:g31247; EMBL:
5; GB:S40856; NID:g31251; EMBL:X62256; GB:S40857; NID:g31252; EMBL:X62257; GB:S40859;
A>Note: sequence extracted from NCBI backbone (NCBIN:109315, NCBIN:109317, NCBIN:1093

```

C:Comment: This apparently silent human homolog of pancreatic elastase I is a single-copy functional protein in some other tissue.

C:Genetics:
A:Gene: GDB:ELA1
A:Map position: 12
C:Keywords: hydrolase; pseudogene; serine proteinase

Query Match 5.4%; Score 115; DB 4; Length 267;
Best Local Similarity 26.9%; Pred. No. 0.042;
Matches 35; Conservative 23; Mismatches 38; Indels 34; Gaps 6;

QY 147 NYPFSTSVKLSGTG-----CTGTLVAEKHVLTAACHIDGKYY--VKGTQKLRVGLPKPK 198
DB 38 SWPSQISLQYRSGSWYHTCGGTLIRQNNWMTAAHCVDYQKTRFVAGDHNL-----89
QY 199 FKDDGGGANDSTSAMPEQMKFQWIRVKRTHVPKGIKGNANDIGMDYDVALLEL-KKPKH 257
DB 90 -----SQNDGTE-----QYVSQKIVVHPYW---NSDINVAAGYDIALRLAQSVTL 132
QY 258 RKFMKIGVSP 267
DB 133 NSYVOLGVLP 142

RESULT 10
TRWV3Y
Trypsin-like proteinase (EC 3.4.21.-) 5G1 precursor - yellow fever mosquito (fragment)
C:Species: Aedes aegypti (yellow fever mosquito)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 18-Jun-1999
C:Accession: S19891
R:Kahnok, S.; Tabak, L.M.; Prosser, D.E.; Downe, A.E.R.; White, B.N.
submitted to the EMBL Data Library, February 1992
A:Description: Isolation, sequencing and characterization of 2 cDNA clones coding for trypsin-like proteinase from the mosquito Aedes aegypti.
A:Reference number: S19890
A:Accession: S19891

A:Molecule type: mRNA
A:Residues: 1-238 <KAL>
A:Cross-references: EMBL:X64363; NID:g5563; PIDN:CAA45715.1; PID:g5564
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; insect midgut; protein digestion; serine proteinase
F:1-11/Domain: signal sequence and propeptide (fragment) #status predicted <SIG>
F:12-238/Product: trypsin-like proteinase 5G1 #status predicted <MAT>
F:12-232/Domain: trypsin homology <TRY>
F:38-54,162-178,189-213/Disulfide bonds: #status predicted
F:53.97,193/Active site: His, Asp, Ser #status predicted

Query Match 5.4%; Score 114.5; DB 1; Length 238;
Best Local Similarity 29.4%; Pred. No. 0.04;
Matches 42; Conservative 22; Mismatches 44; Indels 35; Gaps 7;

QY 149 PFSTSVKLSGTG---VKLSGTGCTGLVAEKHVLTAACHIDGKTYKGTQKLRVGLPKPKKDGGRG 205
DB 24 PQVSLSGVSGSHFCGSLLSERWYTAGCAASGQTNL----QVRIG--SQHASGG-- 75
QY 206 ANDSTSAMPEQMKFQWIRVKRTH--VPKGIKGNANDIGMDYDVALLELKKPKHFKMTG 264
DB 76 -----OLIKVKVNRHPK-----YDEVTTDYDFALLELEETVTFSDSCAP 115
QY 265 VSPPAKQLP---GGRHFSGYDN 284
DB 116 VKLPQKDTVPNEGTCLOVSGWGN 138

RESULT 11
DBHU
Complement factor D (EC 3.4.21.46) precursor [validated] - human (fragment)
N:Alternate names: adipsin; C3 convertase activator
C:Species: Homo sapiens (man)
C:Date: 28-Aug-1985 #sequence_revision 31-Dec-1992 #text_change 08-Dec-2000
C:Accession: A40197; A00936; A60571; S66645
R:White, R.T.; Damm, D.; Hancock, N.; Rosen, B.S.; Lowell, B.B.; Usher, P.; Flier, J.S.;

J. Biol. Chem. 267, 9210-9213, 1992
A:Title: Human adipsin is identical to complement factor D and is expressed at high 1
A:Reference number: A40197; MUID:92250520
A:Accession: A40197
A:Molecule type: mRNA
A:Residues: 1-246 <WHI>
A:Cross-references: GB:M84526
R:Niemann, M.A.; Bhowm, A.S.; Bennett, J.C.; Volanakis, J.E.
Biochemistry 23, 2482-2486, 1984
A:Title: Amino acid sequence of human D of the alternative complement pathway.
A:Reference number: A00936; MUID:85000441
A:Accession: A00936

A:Molecule type: protein
A:Residues: 19-44, 'G', '46-51, 'Q', '53-75, 'TH', '78, 'P', '80-83, 'XXXITIE', '90-172, '86-91, '185-23
A:Note: A few residues were assigned from the previously published sequence of Reid e
R:Miya, T.; Oda, O.; Inagi, R.; Sugiyama, S.; Miyama, A.; Maeda, K.; Nekashima, I.;
Mol. Immunol. 27, 637-644, 1990
A:Title: Molecular and functional identification and purification of complement compo
A:Reference number: A60571; MUID:90370044
A:Accession: A60571

A:Molecule type: protein
A:Residues: 19-20, 'XX', '23-27, 'XX', '30-31, 'XX', '34, 'X', '36-40 <MIY>
R:Balk, N.; Holtkamp, U.; Hoerl, W.H.; Tschesche, H.
FEBS Lett. 371, 300-302, 1995
A:Title: Inhibition of degradation of human polymorphonuclear leukocytes by complem
A:Reference number: S66645; MUID:96013156
A:Accession: S66645

A:Molecule type: preliminary
A:Molecule type: protein
A:Residues: 19-44, 'C', '46-48 <BAL>
C:Comment: Factor D cleaves factor B when the latter is complexed with factor C3b, ac
C:Genetics:
A:Gene: GDB:DF

A:Cross-references: GDB:132645; OMIM:134350
A:Map position: Xpter:Xqter
C:Superfamily: trypsin; trypsin homology
C:Keywords: complement alternate pathway; hydrolase; plasma; serine proteinase
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-246/Product: complement factor D (fragment) #status experimental <MAT>
F:19-241/Domain: trypsin homology <TRY>
F:44-60,141-207,172-188,197-222/Disulfide bonds: #status predicted
F:59,105,201/Active site: His, Asp, Ser #status predicted

Query Match 5.3%; Score 114; DB 1; Length 246;
Best Local Similarity 28.4%; Pred. No. 0.046;
Matches 44; Conservative 20; Mismatches 43; Indels 48; Gaps 9;

QY 149 PFSTSVKLSGTG---CTGTLVAEKHVLTAACHIH---DGKTYVKGTKLRVGLPKPKFDGG 203
DB 31 PYMASVOLNGAHLCAAGLVVAERWVLSAAHCLDADAGKQVVL-----72
QY 204 RGANDSTSAMPEQMKFQWIRVKRTHVPKGIKGNANDIGMDYDVALLELKKPKHFKMKI 263
DB 73 LCAHLSLQPEPCKRLYDVLR-----VPDSDPTIDHLLQLLQSE-----KA 117
QY 264 GVSPPAKQLPGRRIHFSGYDND--RPGNLVYRFCDV 297
DB 118 TLGPAVRPLPWQVR-----DROVAPGTL---CDV 143

RESULT 12
ELRT1

pancreatic elastase (EC 3.4.21.36) I precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 18-Aug-1982 #sequence_revision 18-Aug-1982 #text_change 18-Jun-1999
C:Accession: A00960; A20534
R:MacDonald, R.J.; Swift, G.H.; Quinto, C.; Swain, W.; Pictet, R.L.; Nikovits, W.; Ru
Biochemistry 21, 1453-1463, 1982
A:Title: Primary structure of two distinct rat pancreatic preproelastases determined
A:Reference number: A00960; MUID:82182967
A:Accession: A00960
A:Molecule type: mRNA

```
F;17-26/Domain: activation peptide #status predicted <APT>
F;27-266/Product: elastase I #status experimental <MAT>
F;27-259/Domain: trypsin homology <TRY>
F;56-72,153-200,184-200,210-240/Disulfide bonds: #status experimental
F;71,119,214/Active site: His, Asp, Ser #status experimental

Query Match          5.3%; Score 113; DB 1; Length 266;
Best Local Similarity 27.3%; Pred. No. 0.061;
Matches 36; Conservative 21; Mismatches 41; Indels 34; Gaps 6;

QY 147 NYPFSTSVKLSGTG-----CTGLTVAEKHVLTAACHTHDGKTY--VKGTOKLRVGFELPKP 198
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 37 SWFSQISLQTRSSSWAHTCGGLIRNWNWTAACHVDRELTFRVVVGEHNL----- 88
      ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

QY 199 FKDGGRGANDSTSAMPEQMFKFWIRVRKTRHPVGKWIGNANDIGMDYDALLEL-KKPHK 257
      ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 89 -----NONDQTE-----QQVGVQKIVHPYW---NTDDVAAGYDIALLRLAQSVTL 131
      ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

QY 258 RKEWKIGVSPPA 269
      ::::::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 132 NSYVOLGLVPRA 143
      ::::::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

RESULT 14
TRSMG
trypsin (EC 3.4.21.4) precursor - Streptomyces griseus
C:Species: Streptomyces griseus
C:Date: 24-Apr-1984 #sequence-revision 12-May-1994 #text_change 07-May-1999
C:Accession: JQ1302; A00962
R:Kim, J.C.; Cha, S.H.; Jeong, S.T.; Oh, S.K.; Byun, S.M.
Biochem. Biophys. Res. Commun. 181, 707-713, 1991
A:title: Molecular cloning and nucleotide sequence of Streptomyces griseus trypsin ge
A:Reference number: JQ1302; MUID:92095977
A:Accession: JQ1302
A:Molecule type: DNA
A:Residues: 1-259 <KIM>
A:Cross-references: GB:M64471
A:Experimental source: strain ATCC10137
R:Olafson, R.W.; Jurasek, L.; Carpenter, M.R.; Smillie, L.B.
Biochemistry 14, 1168-1177, 1975
A:title: Amino acid sequence of Streptomyces griseus trypsin. Cyanogen bromide fragme
A:Reference number: A00962; MUID:75127940
A:Accession: A00962
A:Molecule type: protein
A:Residues: 37-95,98-259 <OLA>
R:Read, R.J.; James, M.N.G.
J. Mol. Biol. 200, 523, 1988
A:title: Refined crystal structure of Streptomyces griseus trypsin at 1.7 angstr
A:Reference number: A4574; MUID:88286735
A:Contents: annotation; X-ray crystallography, 1.7 angstroms
A>Note: residues 96-97 modeled as Gly-Ala
C:Genetics:
A:Gene: sprt
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F;1-32/Domain: signal sequence #status predicted <SIG>
F;33-36/Domain: propeptide #status predicted <PRO>
F;37-258/Product: trypsin #status experimental <MAT>
F;37-252/Domain: trypsin homology <TRY>
F;58-74,177-192,204-233/Disulfide bonds: #status experimental
F;73,118,208/Active site: His, Asp, Ser #status experimental

Query Match          5.3%; Score 112.5; DB 1; Length 259;
Best Local Similarity 26.0%; Pred. No. 0.065;
Matches 39; Conservative 21; Mismatches 53; Indels 37; Gaps 5;

QY 148 YPFSTSVKLSGTCTGLTVAEKHVLTAACHTHDGKTYVKGTKLRVGFELPKPFDDGGRCAN 207
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 48 FFP-MVRUSMCGGALYAQDIVLTAAHCY-----SGSGNN 81
      ::::::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

QY 208 DS----TSAMPEOMKFOWIRVRKTRHPVGKWIKGNANDIGMDYDALLELKPKRKPMKIC 264
```

Db 82 TSITATGGVVDLQSSAVKVRSTKVLQA-----PGYNGTKDWALIKLAQPINQPTLKIA 136
QY 265 VSPPAKQLPGGRIFHSFYDNDPGLNLYRRF 294
Db 137 TTYAYNQ---GTFYVAGWGANREGGSGQRY 163

RESULT 15
E85765
hypothetical protein 22592 [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: E85765
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: E85765
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-273 <STO>
A:Cross-references: GB:AE005174; NID:g12515576; PIDN:AG56385.1; GSPDB:GN00145; UMGF:225
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: 22592

Query Match 5.3%; Score 112.5; DB 2; Length 273;
Best Local Similarity 23.3%; Pred. No. 0.069;
Matches 62; Conservative 35; Mismatches 112; Indels 57; Gaps 13;

QY 101 VGIYILSSGSGAQHRDSSGSKSRKROIYGYDSRFSTFGKDFLLNYPFSTSVKLTSG- 159
Db 9 LGAISLTSAFVADKPDVAKSANDEVSTLFFGHDDRPV---NDTTQSPWDVAVGQLETAS 65

QY 160 ---CTGTLVAEKHVLTAHCIHDGKTYVKGTKLRYGFLKPKFKDGR-----GANDS 209
Db 66 GNLCTATLIAPNLALTAGHCL---LTPPKADKAVALRFVSNKGLWRYDIHDIEGRVDP 122

QY 210 TSAMPQOMKFWIRVKTHTVPGWIKGNANDIGMDYDYLLEKPKHKKFKIGVSP-- 267
Db 123 T--LGKRLKAD-----GDGWIVPPA---AAPWDFGLIVLRNPPS-----GITPLP 162

QY 268 -----PAKQLPGGRIFHSFYDNDPGLNLY--YRFCDVKD-ETYDLLYQCCDAQPGA 315
Db 163 LFEGDKAALTAALKAGRKVTQAGYPEDHLDLTLYSHQNCCEVTGWAQTSVASHQCCTLPGD 222

QY 316 SGSGVYVRMWRKQQKWKRIIGIFS 341
Db 223 SGSPMLM----HTDDGQWQ--LIGVQS 242

Search completed: December 2, 2001, 16:17:56
Job time: 191 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:14:40 ; Search time 35.78 Seconds
(without alignments)
246.542 Million cell updates/sec

Title: US-09-072-384-2
Perfect score: 2112
Sequence: 1 MGIPGLLFLFLCAVGO.....IKGNYLDREGDTVFLPGSN 392

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pap.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pap.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pap.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pap.*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pap.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2104	99.6	392	4	US-09-072-384-2
2	2085	98.7	392	4	US-09-072-384-15
3	2044	96.8	383	4	US-08-023-384-18
4	185	7.8	222	1	US-08-090-048-1
5	165	7.8	222	2	US-08-292-550-1
6	165	7.8	222	2	US-07-927-661A-1
7	123	5.8	256	3	US-08-906-769-89
8	123	5.8	256	3	US-08-906-616-89
9	123	5.8	256	4	US-08-817-795-89
10	123	5.8	256	4	US-08-639-075A-89
11	123	5.8	256	4	US-09-012-431-89
12	123	5.8	256	4	US-09-032-215-32
13	123	5.8	256	4	US-09-012-692-89
14	123	5.8	256	4	US-08-906-613-89
15	123	5.8	256	5	PCT-US95-14442A-89
16	121.5	5.8	437	1	US-08-487-037-2
17	118	5.6	241	4	US-08-944-483-59
18	116.5	5.5	437	1	US-08-487-037-3
19	114	5.4	254	1	US-08-330-978-3
20	114	5.4	254	1	US-08-474-042-3
21	114	5.4	254	1	US-08-484-558-3
22	114	5.4	254	1	US-08-774-592-3
23	114	5.4	306	1	US-08-330-978-1
24	114	5.4	306	1	US-08-474-042-1
25	114	5.4	306	1	US-08-484-558-1
26	114	5.4	306	1	US-08-774-592-1
27	114	5.4	448	1	US-08-295-411-3

28	114	5.4	448	2	US-08-955-471-3	Sequence 3, Appli
29	114	5.4	448	5	PCT-US92-10068-1	Sequence 1, Appli
30	114	5.4	448	5	PCT-US92-10242-3	Sequence 3, Appli
31	114	5.4	487	1	US-08-469-486-53	Sequence 53, Appl
32	114	5.4	487	2	US-08-469-658-53	Sequence 53, Appl
33	114	5.4	486	1	US-08-487-037-1	Sequence 1, Appli
34	114	5.4	492	2	US-08-469-486-2	Sequence 2, Appli
35	114	5.4	492	2	US-08-469-658-2	Sequence 2, Appli
36	113.5	5.4	241	1	US-08-330-978-4	Sequence 4, Appli
37	113.5	5.4	241	1	US-08-474-042-4	Sequence 4, Appli
38	113.5	5.4	241	1	US-08-484-558-4	Sequence 4, Appli
39	113.5	5.4	241	1	US-08-774-592-4	Sequence 4, Appli
40	112.5	5.3	223	1	US-08-278-091-13	Sequence 13, Appl
41	112.5	5.3	223	1	US-08-483-859-13	Sequence 13, Appl
42	112.5	5.3	223	1	US-08-472-173-13	Sequence 13, Appl
43	112.5	5.3	223	2	US-08-487-167-13	Sequence 13, Appl
44	112.5	5.3	223	2	US-08-482-816-13	Sequence 13, Appl
45	112.5	5.3	223	2	US-08-296-149-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-09-072-384-2
; Sequence 2, Application US/09072384
; Patent No. 6153420
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,384
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31,648
; REFERENCE/DOCKET NUMBER: 97-16C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 392 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Signal Sequence
; LOCATION: 1..19
; OTHER INFORMATION:
US-09-072-384-2

```
Query Match 99.6%; Score 2104; DB 4; Length 392;
Best Local Similarity 100.0%; Pred. No. 6.6e-219;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGIPGLLFLFLLCVAVGQVSPYSAPWKPTWPAAYRLPVVLPVLPSTLNLAKEPFGAEAKLE 60
DB 1 MAGIPGLLFLFLLCVAVGQVSPYSAPWKPTWPAAYRLPVVLPVLPSTLNLAKEPFGAEAKLE 60
QY 61 VSSCGPQCHKGTPPTYKEAKOYLSYETLYANGSRTEXOVGIYILSSSGDGAAXRDSGS 120
DB 61 VSSCGPQCHKGTPPTYKEAKOYLSYETLYANGSRTEXOVGIYILSSSGDGAAXRDSGS 120
QY 121 SGKSRKRQIYGYDSRFSIFGKDFLLNYPSTSVKLSGTGTLVAEXHVLTAACHIDG 180
DB 121 SGKSRKRQIYGYDSRFSIFGKDFLLNYPSTSVKLSGTGTLVAEXHVLTAACHIDG 180
QY 181 KTVVGTQKLRVGLFKPKFDGGRGANDSTSAMPEOMKFWIRKTHVPKGIKGNAND 240
DB 181 KTVVGTQKLRVGLFKPKFDGGRGANDSTSAMPEOMKFWIRKTHVPKGIKGNAND 240
QY 241 IGMDDYALLELKKPKHKKFKMKGVSPPAKQLPGGRHFGSYDNDPGLNLYVRFCDVKDE 300
DB 241 IGMDDYALLELKKPKHKKFKMKGVSPPAKQLPGGRHFGSYDNDPGLNLYVRFCDVKDE 300
QY 301 TYDLLYQQCDAQPCASGYGVYVWMKRRQOKWERKIIGIFSGHQWMDMNGSPQDFNVAVR 360
DB 301 TYDLLYQQCDAQPCASGYGVYVWMKRRQOKWERKIIGIFSGHQWMDMNGSPQDFNVAVR 360
QY 361 ITPKLYAQICYWIKGNVLDREGDTVFLPGSN 392
DB 361 ITPKLYAQICYWIKGNVLDREGDTVFLPGSN 392

RESULT 2
US-09-072-384-15
; Sequence 15, Application US/09072384
; Patent No. 6153420
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
; TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,384
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31,648
; REFERENCE/DOCKET NUMBER: 97-16C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 392 amino acids
```

```
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Signal Sequence
; LOCATION: 1...19
; OTHER INFORMATION:
US-09-072-384-15

Query Match 98.7%; Score 2085; DB 4; Length 392;
Best Local Similarity 98.2%; Pred. No. 7.5e-217;
Matches 395; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MAGIPGLLFLFLLCVAVGQVSPYSAPWKPTWPAAYRLPVVLPVLPSTLNLAKEPFGAEAKLE 60
DB 1 MAGIPGLLFLFLLCVAVGQVSPYSAPWKPTWPAAYRLPVVLPVLPSTLNLAKEPFGAEAKLE 60
QY 61 VSSCGPQCHKGTPPTYKEAKOYLSYETLYANGSRTEXOVGIYILSSSGDGAAXRDSGS 120
DB 61 VSSCGPQCHKGTPPTYKEAKOYLSYETLYANGSRTEXOVGIYILSSSGDGAAXRDSGS 120
QY 121 SGKSRKRQIYGYDSRFSIFGKDFLLNYPSTSVKLSGTGTLVAEXHVLTAACHIDG 180
DB 121 SGKSRKRQIYGYDSRFSIFGKDFLLNYPSTSVKLSGTGTLVAEXHVLTAACHIDG 180
QY 181 KTVVGTQKLRVGLFKPKFDGGRGANDSTSAMPEOMKFWIRKTHVPKGIKGNAND 240
DB 181 KTVVGTQKLRVGLFKPKFDGGRGANDSTSAMPEOMKFWIRKTHVPKGIKGNAND 240
QY 241 IGMDDYALLELKKPKHKKFKMKGVSPPAKQLPGGRHFGSYDNDPGLNLYVRFCDVKDE 300
DB 241 IGMDDYALLELKKPKHKKFKMKGVSPPAKQLPGGRHFGSYDNDPGLNLYVRFCDVKDE 300
QY 301 TYDLLYQQCDAQPCASGYGVYVWMKRRQOKWERKIIGIFSGHQWMDMNGSPQDFNVAVR 360
DB 301 TYDLLYQQCDAQPCASGYGVYVWMKRRQOKWERKIIGIFSGHQWMDMNGSPQDFNVAVR 360
QY 361 ITPKLYAQICYWIKGNVLDREGDTVFLPGSN 392
DB 361 ITPKLYAQICYWIKGNVLDREGDTVFLPGSN 392

RESULT 3
US-09-072-384-18
; Sequence 18, Application US/09072384
; Patent No. 6153420
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
; TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,384
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:15:33 ; Search time 35.78 seconds
(without alignments)
246.542 Million cell updates/sec

Title: US-09-072-384-15

Perfect score: 2131

Sequence: 1 MAGIPGLFLFLFLCAVQ.....IKGNLYDCRGDTVPFGSN 392

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA.*

- 1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2131	100.0	392	4	US-09-072-384-15
2	2085	97.8	392	4	US-09-072-384-2
3	2080	97.6	383	4	US-09-072-384-18
4	177	8.3	222	1	US-08-090-048-1
5	177	8.3	222	2	US-08-292-550-1
6	177	8.3	222	2	US-07-927-661A-1
7	122	5.7	256	3	US-08-906-769-89
8	122	5.7	256	3	US-08-906-616-89
9	122	5.7	256	3	US-08-817-795-89
10	122	5.7	256	4	US-08-639-075A-89
11	122	5.7	256	4	US-09-012-431-89
12	122	5.7	256	4	US-09-032-215-32
13	122	5.7	256	4	US-09-012-692-89
14	122	5.7	256	4	US-08-906-613-89
15	122	5.7	256	5	PCT-US95-14442A-89
16	117	5.5	437	1	US-08-487-037-2
17	115.5	5.4	241	4	US-08-944-483-59
18	114	5.3	228	4	US-08-944-483-44
19	114	5.3	253	6	5223425-8
20	113	5.3	238	6	5223425-5
21	113	5.3	250	6	5223425-4
22	112.5	5.3	223	1	US-08-278-091-13
23	112.5	5.3	223	1	US-08-483-859-13
24	112.5	5.3	223	1	US-08-472-173-13
25	112.5	5.3	223	2	US-08-487-167-13
26	112.5	5.3	223	2	US-08-482-816-13
27	112.5	5.3	223	2	US-08-296-149-13

28 112.5 5.3 223 2 US-08-801-499-13 Sequence 13, Appl
29 112.5 5.3 223 2 US-08-615-271-13 Sequence 13, Appl
30 112.5 5.3 223 3 US-09-074-660-13 Sequence 13, Appl
31 112.5 5.3 223 3 US-09-074-659-13 Sequence 13, Appl
32 112.5 5.3 223 3 US-09-106-468-13 Sequence 13, Appl
33 112.5 5.3 223 4 US-09-106-466A-13 Sequence 13, Appl
34 112.5 5.3 223 4 US-09-106-467-13 Sequence 13, Appl
35 111.5 5.2 241 1 US-08-330-978-4 Sequence 4, Appl
36 111.5 5.2 241 1 US-08-474-042-4 Sequence 4, Appl
37 111.5 5.2 241 1 US-08-484-558-4 Sequence 4, Appl
38 111.5 5.2 241 1 US-08-774-592-4 Sequence 4, Appl
39 111.5 5.2 254 1 US-08-330-978-3 Sequence 3, Appl
40 111.5 5.2 254 1 US-08-474-042-3 Sequence 3, Appl
41 111.5 5.2 254 1 US-08-484-558-3 Sequence 3, Appl
42 111.5 5.2 254 1 US-08-774-592-3 Sequence 3, Appl
43 111.5 5.2 306 1 US-08-330-978-1 Sequence 1, Appl
44 111.5 5.2 306 1 US-08-474-042-1 Sequence 1, Appl
45 111.5 5.2 306 1 US-08-484-558-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-072-384-15
; Sequence 15, Application US/09072384
; Patent No. 6153420
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
; TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,384
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31,648
; REFERENCE/DOCKET NUMBER: 97-16C1
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 392 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Signal Sequence
; LOCATION: 1..19
; OTHER INFORMATION:
US-09-072-384-15

```
Query Match 100.0%; Score 2131; DB 4; Length 392;
Best Local Similarity 100.0%; Pred. No. 1.1e-227;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGIPGLLFLFFLLCAVGVSPYSAPWKPWPAYRLPVVLPVLPSTLNLAKEPDKFAEAKLE 60
DB 1 MAGIPGLLFLFFLLCAVGVSPYSAPWKPWPAYRLPVVLPVLPSTLNLAKEPDKFAEAKLE 60
QY 61 VSSSCGPQCHKGTPPTYEAKOYLSYETLYANGSRTEQVGIYILSSSGDGAQHRDGS 120
DB 61 VSSSCGPQCHKGTPPTYEAKOYLSYETLYANGSRTEQVGIYILSSSGDGAQHRDGS 120
QY 121 SGSRKRQIYGYDSRFSIFGKDFLLNYPSTSVKLSGTCTGLVAEKHVLTAACHIDG 180
DB 121 SGSRKRQIYGYDSRFSIFGKDFLLNYPSTSVKLSGTCTGLVAEKHVLTAACHIDG 180
QY 181 KTVVKTOKLRVGLFKPKFKDGGRGANDSTSAMPEQMKFQWIRKTHVPKGIKGNAND 240
DB 181 KTVVKTOKLRVGLFKPKFKDGGRGANDSTSAMPEQMKFQWIRKTHVPKGIKGNAND 240
QY 241 IGMDDYALLELAKKPKRKFMTIGVSPPAKQLPGGRHFSGYDNDRPNLVYRFDVKDE 300
DB 241 IGMDDYALLELAKKPKRKFMTIGVSPPAKQLPGGRHFSGYDNDRPNLVYRFDVKDE 300
QY 301 TYDLLYQQCDAQPGASGVYVVRMKRQKQKWERKIIGIFSGHQWDMNGSPQDFNVAVR 360
DB 301 TYDLLYQQCDAQPGASGVYVVRMKRQKQKWERKIIGIFSGHQWDMNGSPQDFNVAVR 360
QY 361 ITPLKVAQICYWIKGNLYDCREGDTVPFGSN 392
DB 361 ITPLKVAQICYWIKGNLYDCREGDTVPFGSN 392

RESULT 2
US-09-072-384-2
; Sequence 2, Application US/09072384
; Patent No. 6153420
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
; TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: zymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,384
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31,648
; REFERENCE/DOCKET NUMBER: 97-16C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 392 amino acids
```

```
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
FEATURE:
NAME/KEY: Signal Sequence
LOCATION: 1...19
OTHER INFORMATION:
US-09-072-384-2

Query Match 97.8%; Score 2085; DB 4; Length 392;
Best Local Similarity 98.2%; Pred. No. 1.5e-222;
Matches 385; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MAGIPGLLFLFFLLCAVGVSPYSAPWKPWPAYRLPVVLPVLPSTLNLAKEPDKFAEAKLE 60
DB 1 MAGIPGLLFLFFLLCAVGVSPYSAPWKPWPAYRLPVVLPVLPSTLNLAKEPDKFAEAKLE 60
QY 61 VSSSCGPQCHKGTPPTYEAKOYLSYETLYANGSRTEQVGIYILSSSGDGAQHRDGS 120
DB 61 VSSSCGPQCHKGTPPTYEAKOYLSYETLYANGSRTEQVGIYILSSSGDGAQHRDGS 120
QY 121 SGSRKRQIYGYDSRFSIFGKDFLLNYPSTSVKLSGTCTGLVAEKHVLTAACHIDG 180
DB 121 SGSRKRQIYGYDSRFSIFGKDFLLNYPSTSVKLSGTCTGLVAEKHVLTAACHIDG 180
QY 181 KTVVKTOKLRVGLFKPKFKDGGRGANDSTSAMPEQMKFQWIRKTHVPKGIKGNAND 240
DB 181 KTVVKTOKLRVGLFKPKFKDGGRGANDSTSAMPEQMKFQWIRKTHVPKGIKGNAND 240
QY 241 IGMDDYALLELAKKPKRKFMTIGVSPPAKQLPGGRHFSGYDNDRPNLVYRFDVKDE 300
DB 241 IGMDDYALLELAKKPKRKFMTIGVSPPAKQLPGGRHFSGYDNDRPNLVYRFDVKDE 300
QY 301 TYDLLYQQCDAQPGASGVYVVRMKRQKQKWERKIIGIFSGHQWDMNGSPQDFNVAVR 360
DB 301 TYDLLYQQCDAQPGASGVYVVRMKRQKQKWERKIIGIFSGHQWDMNGSPQDFNVAVR 360
QY 361 ITPLKVAQICYWIKGNLYDCREGDTVPFGSN 392
DB 361 ITPLKVAQICYWIKGNLYDCREGDTVPFGSN 392

RESULT 3
US-09-072-384-18
; Sequence 18, Application US/09072384
; Patent No. 6153420
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
; TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: zymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,384
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:14:40 ; Search time 64.47 seconds
(without alignments)
450.391 Million cell updates/sec

Title: US-09-072-384-2
Perfect score: 2112
Sequence: 1 MAGIPGLLFLFLCAVGO.....IKGNYLDCRGDTVFLPGSN 392

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SID22/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	2104	99.6	392	22	Human Zsig13 varia
2	2085	98.7	392	22	Human Zsig13 varia
3	2044	96.8	383	20	Human transmembran
4	2044	96.8	383	20	WO9927094 Seq ID 1
5	2044	96.8	383	20	Amino acid sequenc
6	2044	96.8	383	21	Protein encoded by
7	2044	96.8	383	21	Protein encoded by
8	2044	96.8	383	21	Human TANGO 186 pr
9	2044	96.8	383	21	Human signal pepti
10	2044	96.8	383	21	A bone marrow secr
11	2044	96.8	383	22	Human PRO307 prote

12	2044	96.8	383	22	Human Zsig13 varia
13	2037	95.4	383	22	Human membrane or
14	1873.5	88.7	375	21	Protein encoded by
15	1873.5	88.7	413	21	Protein encoded by
16	1868.5	88.5	382	21	Marine TANGO 186 p
17	1028	48.7	413	20	Human PRO1057 prot
18	1028	48.7	413	21	Human nn320-2 secr
19	1028	48.7	413	21	Human PRO1057 (UNQ
20	190.5	9.0	314	22	Bacillus lichenifo
21	185	8.8	316	13	Bacillus lichenifo
22	184	8.7	316	22	Protease BLase. S
23	172	8.1	222	22	Bacillus lichenifo
24	171	8.1	222	22	Bacillus lichenifo
25	171	8.1	222	22	Bacillus lichenifo
26	171	8.1	222	22	Bacillus lichenifo
27	171	8.1	222	22	Bacillus lichenifo
28	171	8.1	222	22	Bacillus lichenifo
29	169	8.0	222	22	Bacillus lichenifo
30	169	8.0	222	22	Bacillus lichenifo
31	169	8.0	222	22	Bacillus lichenifo
32	169	8.0	222	22	Bacillus lichenifo
33	168	8.0	222	22	Bacillus lichenifo
34	168	8.0	222	22	Bacillus lichenifo
35	168	8.0	222	22	Bacillus lichenifo
36	168	8.0	222	22	Bacillus lichenifo
37	167	7.9	222	22	Bacillus lichenifo
38	167	7.9	222	22	Bacillus lichenifo
39	167	7.9	222	22	Bacillus lichenifo
40	167	7.9	222	22	Bacillus lichenifo
41	167	7.9	222	22	Bacillus lichenifo
42	167	7.9	222	22	Bacillus lichenifo
43	167	7.9	222	22	Bacillus lichenifo
44	167	7.9	222	22	Bacillus lichenifo
45	167	7.9	222	22	Bacillus lichenifo

ALIGNMENTS

RESULT 1
AAB48972
ID AAB48972 standard; Protein; 392 AA.
XX AAB48972;
AC AAB48972;
DT 27-MAR-2001 (first entry)
XX
DE Human Zsig13 variant #1, SEQ ID NO:2.
XX
KW Human Zsig13; serine protease; chromosome 11q22.1; elastase homologue;
KW glutamyl endopeptidase homologue; factor X homologue; trypsin homologue;
KW trypsinogen homologue; mast cell protease homologue;
KW collagenase homologue; protein degradation; food processing; brewing;
KW alcohol production; laundry detergent component.
XX
OS Homo sapiens.
XX
PN US6153420-A.
PD 28-NOV-2000.
XX
PF 04-MAY-1998; 98US-0072384.
XX
PR 24-APR-1997; 97US-0044185.
PR 17-APR-1998; 98US-0062142.
XX
PA (ZYMO) ZYMOGENETICS INC.
XX
PI Sheppard PO; - inventor
XX
DR WPI; 2001-060090/07.
DR N-PSDB; AAC91782.
XX

PT New isolated serine protease (designated Zsig13), useful in industrial
 PT processes to degrade unwanted proteins or alter the characteristics of
 PT protein-containing composition, as well as in industrial applications
 PT (e.g. brewing)

XX Claim 1; Column 25-28; 26pp; English.

XX The invention relates to human Zsig13 proteins (AAB48972-B48974), and
 CC to DNA encoding them (AAC91782-C91784). The invention also relates to
 CC expression vectors and host cells comprising a human Zsig13 DNA, and the
 CC recombinant production of a human Zsig13 protein or its precursor.
 CC Zsig13 is a serine protease, and has significant homology to Bacillus
 CC licheniformis glutamyl endopeptidase, human clotting factor X, human
 CC elastase, rat mast cell protease, Streptomyces griseus trypsin, bovine
 CC trypsinogen, and Hypoderma lineatum collagenase. The gene encoding human
 CC Zsig13 is located on chromosome 11q22.1. Zsig13 is useful in industrial
 CC processes to degrade unwanted proteins or alter the characteristics of
 CC protein-containing compositions. It may also be used in industrial
 CC applications in which proteases are utilised, including food processing,
 CC brewing and alcohol production, and as a component of a laundry
 CC detergent. The present sequence represents a human Zsig13 variant.

XX Sequence 392 AA;

Query Match 99.6%; Score 2104; DB 22; Length 392;
 Best Local Similarity 100.0%; Pred. No. 3.9e-147;
 Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGIPGLLLFLLCAVGOVSPYSAPKPTWPAYRLPVVLPQSTLNLAKEPFGAEAKLE 60
 DB 1 magipglillfllcavgvspysapkwptwpayrlpvvlpqstlnlakpfgaeakle 60
 QY 61 VSSCGPQCHKGTPPTYKEAKOYLSYETLYANGSRTEXOVGIYILSSGDGAXXRDSGS 120
 DB 61 vsscgpqchkgtptpykeakylysetlyangsrtevxgvilylssgdgaxxrdsgs 120
 QY 121 SGKRRRKQIYGYDSRFSIFGKDFLLNYPFSTSVKLTGCTGLVAEXHVLTAACHIDHG 180
 DB 121 sgkrrrkqiygydsrfsifgkdfllnypfstsvklstgctglvaexhvltaachidhg 180
 QY 181 KTYVKGTKLRVGLFKPKFKDGGGANDSTSAMPEQMKFQWIRKRVHPKGIKGNAND 240
 DB 181 ktyvkgtklrvglfkpkfkdggrgandstsampeqmkfgwlrkvrthvphkgnand 240
 QY 241 IGMDDYVALLLELKKPKRKFEMKIGVSPAKQLPGGRTHFSGYDNDRPNLVYRCDVKDE 300
 DB 241 igmddyvallylelkkpkrrkfemkigvspakqlpggrthfsgydnrdpnlvyrfdvkd 300
 QY 301 TYDLLYQCCDAQPGASGYGVYVRMWRKQOQKWERKIIGIFSGHQWDMNGSPQDFNVAVR 360
 DB 301 tydlllyqcdapqasgygyvvrwmwrkqqkwerkiigifsghqwdmngspqdfnvavr 360
 QY 361 ITPLYKAIQICWIKGNVLDREGDTVFLPGSN 392
 DB 361 itplykaiqicwvknvldcregdtvflpgsn 392

RESULT 2

ID AAB48973 standard; Protein; 392 AA.

XX AAB48973;

XX AAB48973;

XX 27-MAR-2001 (first entry)

XX Human Zsig13 variant #2, SEQ ID NO:15.

XX Human Zsig13;

XX Human Zsig13; serine protease; chromosome 11q22.1; elastase homologue;

XX glutamyl endopeptidase homologue; factor X homologue; trypsin homologue;

XX trypsinogen homologue; mast cell protease homologue;

XX collagenase homologue; protein degradation; food processing; brewing;

XX alcohol production; laundry detergent component.

XX Homo sapiens.
 XX US6153420-A.
 XX 28-NOV-2000.
 XX 04-MAY-1998; 98US-0072384.
 XX 24-APR-1997; 97US-0044185.
 XX 17-APR-1998; 98US-0062142.

XX (ZYMO) ZYMOGENETICS INC.

XX Sheppard PQ;

XX WPI; 2001-060090/07.

XX N-PSDB; AAC91783.

XX New isolated serine protease (designated Zsig13), useful in industrial
 PT processes to degrade unwanted proteins or alter the characteristics of
 PT protein-containing composition, as well as in industrial applications
 PT (e.g. brewing)

XX Claim 1; Column 35-38; 26pp; English.

XX The invention relates to human Zsig13 proteins (AAB48972-B48974), and
 CC to DNA encoding them (AAC91782-C91784). The invention also relates to
 CC expression vectors and host cells comprising a human Zsig13 DNA, and the
 CC recombinant production of a human Zsig13 protein or its precursor.
 CC Zsig13 is a serine protease, and has significant homology to Bacillus
 CC licheniformis glutamyl endopeptidase, human clotting factor X, human
 CC elastase, rat mast cell protease, Streptomyces griseus trypsin, bovine
 CC trypsinogen, and Hypoderma lineatum collagenase. The gene encoding human
 CC Zsig13 is located on chromosome 11q22.1. Zsig13 is useful in industrial
 CC processes to degrade unwanted proteins or alter the characteristics of
 CC protein-containing compositions. It may also be used in industrial
 CC applications in which proteases are utilised, including food processing,
 CC brewing and alcohol production, and as a component of a laundry
 CC detergent. The present sequence represents a human Zsig13 variant.

XX Sequence 392 AA;

Query Match 98.7%; Score 2085; DB 22; Length 392;
 Best Local Similarity 98.2%; Pred. No. 9.7e-146;
 Matches 385; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MAGIPGLLLFLLCAVGOVSPYSAPKPTWPAYRLPVVLPQSTLNLAKEPFGAEAKLE 60
 DB 1 magipglillfllcavgvspysapkwptwpayrlpvvlpqstlnlakpfgaeakle 60
 QY 61 VSSCGPQCHKGTPPTYKEAKOYLSYETLYANGSRTEXOVGIYILSSGDGAXXRDSGS 120
 DB 61 vsscgpqchkgtptpykeakylysetlyangsrtevxgvilylssgdgaxxrdsgs 120
 QY 121 SGKRRRKQIYGYDSRFSIFGKDFLLNYPFSTSVKLTGCTGLVAEXHVLTAACHIDHG 180
 DB 121 sgkrrrkqiygydsrfsifgkdfllnypfstsvklstgctglvaexhvltaachidhg 180
 QY 181 KTYVKGTKLRVGLFKPKFKDGGGANDSTSAMPEQMKFQWIRKRVHPKGIKGNAND 240
 DB 181 ktyvkgtklrvglfkpkfkdggrgandstsampeqmkfgwlrkvrthvphkgnand 240
 QY 241 IGMDDYVALLLELKKPKRKFEMKIGVSPAKQLPGGRTHFSGYDNDRPNLVYRCDVKDE 300
 DB 241 igmddyvallylelkkpkrrkfemkigvspakqlpggrthfsgydnrdpnlvyrfdvkd 300
 QY 301 TYDLLYQCCDAQPGASGYGVYVRMWRKQOQKWERKIIGIFSGHQWDMNGSPQDFNVAVR 360
 DB 301 tydlllyqcdapqasgygyvvrwmwrkqqkwerkiigifsghqwdmngspqdfnvavr 360
 QY 361 ITPLYKAIQICWIKGNVLDREGDTVFLPGSN 392

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: December 2, 2001, 16:15:41 ; Search time 32.13 seconds
(without alignments)
447.327 Million cell updates/sec

Title: us-09-072-384-2

Perfect score: 2112

Sequence: 1 MAGIPGLLFLFLCAVQ.....IKGNLYDCREGDTVFLPGSN 392

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	185	8.8	316	1 GSEP_BACLI	P80057 bacillus li
2	130.5	6.2	313	1 MPR_BACSU	P3790 bacillus su
3	122.5	5.8	490	1 FA10_RABIT	O19045 oryctolagus
4	119.5	5.7	376	1 FA10_TROCA	P81428 troglodectis
5	118	5.6	266	1 EL1_BOVIN	P28153 bos taurus
6	118	5.6	269	1 EL2A_HUMAN	P08217 homo sapien
7	114	5.4	488	1 FA10_HUMAN	P00742 homo sapien
8	114	5.4	492	1 FA10_BOVIN	P00743 bos taurus
9	113	5.4	269	1 EL2_PIG	P08419 sus scrofa
10	113	5.4	271	1 EL2_MOUSE	P05208 mus musculus
11	112.5	5.3	259	1 TRYP_STRGR	P00775 streptomyce
12	112	5.3	266	1 EL1_PIG	P00772 sus scrofa
13	112	5.3	266	1 EL1_RAT	P00773 rattus norv
14	111.5	5.3	238	1 TRYS_AEDAE	P29787 aedes aegypt
15	111.5	5.3	686	1 MAS2_HUMAN	O00187 homo sapien
16	110	5.2	269	1 EL2B_HUMAN	P08218 homo sapien
17	109.5	5.2	875	1 NETR_HUMAN	P56730 homo sapien
18	109	5.2	253	1 CFAD_HUMAN	P00746 homo sapien
19	109	5.2	761	1 NETR_MOUSE	O08762 mus musculus
20	107	5.1	269	1 EL2_BOVIN	Q29461 bos taurus
21	106	5.0	271	1 EL2_RAT	P00774 rattus norv
22	105	5.0	603	1 CFAL_MOUSE	Q61129 mus musculus
23	104.5	4.9	268	1 TRYP_STRGA	Q54179 streptomyce
24	104.5	4.9	273	1 YGCD_ECOLI	P76176 escherichia
25	102	4.8	786	1 STUB_DROME	O05319 drosophila
26	101.5	4.8	583	1 CFAL_HUMAN	P05156 homo sapien
27	99.5	4.7	281	1 TRY2_DROBR	P54630 drosophila
28	99.5	4.7	454	1 TMS3_HUMAN	P57727 homo sapien
29	99	4.7	274	1 TRY5_ANOGA	P35039 anopheles g
30	99	4.7	604	1 CFAL_RAT	Q94433 rattus norv
31	97.5	4.6	259	1 DEF3_DERFA	P49275 dermatophag
32	96	4.5	407	1 FA7_BOVIN	P22457 bos taurus
33	95.5	4.5	261	1 DER3_DERPT	P39675 dermatophag

RESULT 1

ID	GSEP_BACLI	STANDARD:	PRT:	316 AA.
DT	01-MAY-1992 (Rel. 22, Created)			
DT	01-OCT-1993 (Rel. 27, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	GLUTAMYL ENDOPEPTIDASE PRECURSOR (EC 3.4.21.19) (GLUTAMATE SPECIFIC			
DE	ENDOPEPTIDASE) (GSE).			
GN	BLASE.			
OS	Bacillus licheniformis.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1402;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	STRAIN=ATCC 14580;			
RX	MEDLINE=93054737; PubMed=1429718;			
RA	Okamoto S., Kikuchi N., Kitadokoro K., Fujiwara T., Nakamura E.,			
RA	Okamoto H., Shin M., Tamaki M., Teraoka H., Tsuzuki H., Yoshida N.;			
RT	"Purification, characterization, cloning, and expression of a			
RT	putamic acid-specific protease from Bacillus licheniformis ATCC			
RT	14580.";			
RL	J. Biol. Chem. 267:23782-23788(1992).			
RL	[2]			
RP	SEQUENCE OF 95-316.			
RX	MEDLINE=92155199; PubMed=1346764;			
RA	Svensden I., Bredam K.;			
RT	"Isolation and amino acid sequence of a glutamic acid specific			
RT	endopeptidase from Bacillus licheniformis.";			
RL	Eur. J. Biochem. 204:165-171(1992).			
CC	-!- FUNCTION: SPECIFIC FOR HYDROLYSIS OF PEPTIDES BONDS ON THE			
CC	CARBOXYL SIDE OF ACIDIC AMINO ACID RESIDUES, WITH A STRONG			
CC	PREFERENCE FOR GLU.			
CC	-!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ASP-I-XAA, GLU-I-XAA.			
CC	-!- SUBCELLULAR LOCATION: SECRETED.			
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE			
CC	V8 FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: D10060; BAA00949.1; -.			
DR	PIR: S23078; S23078.			
DR	PIR: A45134; A45134.			
DR	MEROPS: S01.271; -.			
DR	InterPro: IPR00126; Ser_proteas_v8.			
DR	InterPro: IPR001254; Trypsin.			
DR	Pfam: PF00089; trypsin; 1.			
DR	PRINTS: PR000839; V8PROTEASE.			
DR	SMART: SM00020; TRYP_SPC; 1.			

Q9Y5Y6 homo sapien
P56677 mus musculus
P00741 bos taurus
Q01513 podospora a
P19540 canis famli
O97370 euroglyphus
Q99895 homo sapien
P05049 drosophila
P24664 saccharopol
P52905 drosophila
P08897 hypoderma l
P21844 mus musculus

ALIGNMENTS

```
DR PROSITE; PS00672; V8_HIS; 1.
DR PROSITE; PS00673; V8_SER; 1.
KW Hydrolase; Serine protease; Signal.
FT SIGNAL 1 ?
FT PROPEP 2 94
FT CHAIN 95 316 GLUTAMYL ENDOPEPTIDASE.
FT ACT_SITE 141 141 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 261 261 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 126 142
FT DISULFID 275 279
SQ SEQUENCE 316 AA; 96D7552CB7089B09 CRC64;

Query Match 8.8%; Score 185; DB 1; Length 316;
Best Local Similarity 23.2%; Pred. No. 6.6e-09;
Matches 77; Conservative 46; Mismatches 139; Indels 70; Gaps 16;

QY 63 SSCGPOCHKGTPL---PTYEAKQYLYETLYANGSTEXQVGIYILSSSGDGAXXRDSG 119
: | | | | | : : : : : : : : : : : : : : : : : : : : : :
Db 28 AQAAPSPH--TPVSSDPSYK-AETSVYDP-----NIKSDQYGLYKAFGTG---KVNE 76
: | | | | | : : : : : : : : : : : : : : : : : : : : : :

QY 120 SSGKSRK-----RQIVGDSRFSGKDFLLNYPFSTSVKLST---GCTGTILVAE 167
: | | | | | : : : : : : : : : : : : : : : : : : : : : :
Db 77 TKEAKRSKAPAKYISKVIGSDRTRVN---TTAYPIRAIVHSSIGSCCTGWMIGP 133
: | | | | | : : : : : : : : : : : : : : : : : : : : : :

QY 168 XHVLTAACHIDGKT-YVGTQKLRVGLPKPKFDGGRGANDSTSAMPEQMKFQWIRKVR 226
: | | | | | : : : : : : : : : : : : : : : : : : : : : :
Db 134 KTVATAGHCIVDTSSGSFAGTAVSPG-----RNGTFS-----YPIGSKVSTR 175
: | | | | | : : : : : : : : : : : : : : : : : : : : : :

QY 227 THVPKGIKNANDIGMDYVALLEKKPKRPMKTCVSPPAKQLPGRHIFSGYDNR 286
: | | | | | : : : : : : : : : : : : : : : : : : : : : :
Db 176 YFIPSGWRSGNTN-----YDGAIEELSEPIGNTVGYFGSYTTSSLVGTGTTVTSIGYPGDK 230
: | | | | | : : : : : : : : : : : : : : : : : : : : : :

QY 287 PGNLVYRFCD-VKDERYDLYQCDQAQPGASGVGVVVRMWRKQOQKWKRII---GIF 340
: | | | | | : : : : : : : : : : : : : : : : : : : : : :
Db 231 TAGTQWHSFGPIAISEYKQLQYAM-DYGGQSGSPVEQSSRTNCSPCSLAVHTNGVY 289
: | | | | | : : : : : : : : : : : : : : : : : : : : : :

QY 341 SGHGWDMGSPQDNVAVRITPLKYAICVW 372
: | | | | | : : : : : : : : : : : : : : : : : : : : : :
Db 290 GG-----SSYNGRTTRITKEVFONLTNW 311
: | | | | | : : : : : : : : : : : : : : : : : : : : : :

RESULT 2
MPR_BACSU STANDARD; PRT; 313 AA.
AC P39730;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EXTRACELLULAR METALLOPROTEASE PRECURSOR (EC 3.4.21.-).
GN MPR.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=GP241;
RX MEDLINE=90130256; PubMed=2105291;
RA Sloma A., Rudolph C.F., Rufo G.A. Jr., Sullivan B.J., Theriault K.A.,
RA Ally D., Pero J.;
RT "Gene encoding a novel extracellular metalloprotease in Bacillus
RT subtilis.";
RL J. Bacteriol. 172:1024-1029(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Haga K., Liu H., Yasumoto K., Takahashi H., Yoshikawa H.;
RT "Sequence analysis of the 70kb region between 17 and 23 degree of the
RT Bacillus subtilis chromosome.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-68 FROM N.A.
```

```
RX MEDLINE=89108019; PubMed=3145906;
RA Smith H., de Jong A., Bron S., Venema G.;
RT "Characterization of signal-sequence-coding regions selected from the
RT Bacillus subtilis chromosome.";
RL Gene 70:351-361(1988).
RN [4]
RN SEQUENCE OF 1-10 FROM N.A.
RC STRAIN=168;
RX MEDLINE=96118702; PubMed=7496533;
RA Saxild H.H., Jacobsen J.H., Nygaard P.;
RT "Functional analysis of the Bacillus subtilis purf gene encoding
RT formate-dependent glycinamide ribonucleotide transformylase.";
RL Microbiology 141:2211-2218(1995).
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE
CC V8 FAMILY.
CC -1- CAUTION: CALLED "METALLOPROTEASE", BUT CLEARLY BELONGS TO THE S2
CC FAMILY OF SERINE PROTEASES.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; L10505; AAA22604.1; -
DR EMBL; AB006424; BAA33121.1; -
DR EMBL; M22916; AAA22832.1; -
DR EMBL; Z99105; CAB12018.1; -
DR PIR; A35122; A35122.
DR HSP; P00756; 1SGF.
DR MEROPS; S01.272; -
DR Subtilist; BGI0690; mpr.
DR InterPro; IPR000126; Ser_proteas_V8.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00839; V8PROTEASE.
DR SMART; SM0020; Tryp_SPC; 1.
DR PROSITE; PS00672; V8_HIS; 1.
DR PROSITE; PS00673; V8_SER; 1.
KW Hydrolase; Serine protease; Signal; Zymogen; Complete proteome.
FT SIGNAL 1 34
FT PROPEP 35 93 EXTRACELLULAR METALLOPROTEASE.
FT CHAIN 94 313 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 146 146 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 131 147 BY SIMILARITY.
FT CONFLICT 61 68 QVSAPYEG -> PLESTAQA (IN REF. 3).
SQ SEQUENCE 313 AA; 33842 MW; D41788ED652AE94 CRC64;

Query Match 6.2%; Score 130.5; DB 1; Length 313;
Best Local Similarity 22.6%; Pred. No. 0.00038;
Matches 87; Conservative 38; Mismatches 139; Indels 121; Gaps 19;

QY 28 WKTWPAYRLPVVLQSTLNLAQDFCAEAKLEVS-----SSCGPOCHKGTPLPTYEAK 82
: | | | | | : : : : : : : : : : : : : : : : : : : : : :
Db 7 FRQWFAYLVLCLALA---AAVSGVPAKAENPQTSVNTGKEA-DATKNQTSKADQ 61
: | | | | | : : : : : : : : : : : : : : : : : : : : : :

QY 83 QYLSYE-----TLYANGSRTEQVGIYILSSSGDGAXXRDSGSKSRKRQIYGYDS 135
: | | | | | : : : : : : : : : : : : : : : : : : : : : :
Db 62 VSAPYEGTGKTSKSLY--GGQTELEKNIQTLQPS-----SIIGTDE 100
: | | | | | : : : : : : : : : : : : : : : : : : : : : :

QY 136 RFSIFGKDFLLNYPFSTSVKLST-----GCTGTILVAEXHVLTAACHIH----- 178
: | | | | | : : : : : : : : : : : : : : : : : : : : : :
Db 101 RTRI---SSTSPFYRTVLQSLKYPNTSSTYGTGLVNPNTVVTAGHCVYSDHGWAS 157
: | | | | | : : : : : : : : : : : : : : : : : : : : : :

QY 179 -----DCKTVKGTQKLRVGLPKPKFDGGRGANDSTSAMPEQMKFQWIRKTHVP 230
: | | | | | : : : : : : : : : : : : : : : : : : : : : :
Db 158 TITAAPCRNGSSYPYGYI-----SGTMFYSVK-GWTESKDNYD 195
: | | | | | : : : : : : : : : : : : : : : : : : : : : :
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:14:45 ; Search time 48.03 Seconds
(without alignments)
621.704 Million cell updates/sec

Title: US-09-072-384-2

Perfect score: 2112

Sequence: 1 MAGIPGLLFLFLCAVGG.....IKGYLDCRGDTVFLPGSN 392

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	185	8.8	316	2 A45134	endopeptidase (EC
2	130.5	6.2	313	2 A35122	metalloproteinase
3	125	5.9	482	1 EXRT	coagulation factor
4	118	5.6	269	2 B26823	pancreatic elastas
5	114	5.4	258	2 S70439	pancreatic elastasi
6	114	5.4	267	4 A36615	probable pancreati
7	114	5.4	488	1 EXHU	coagulation factor
8	114	5.4	492	1 EXBO	coagulation factor
9	113	5.4	269	2 A26823	pancreatic elastasi
10	113	5.4	271	2 A25528	pancreatic elastasi
11	112.5	5.3	259	1 TRSMG	trypsin (EC 3.4.21
12	112	5.3	266	1 ELRT1	pancreatic elastasi
13	112	5.3	266	1 ELPG	pancreatic elastasi
14	112	5.3	522	2 T29767	pancreatic elastasi
15	111.5	5.3	238	1 TRWV5Y	hypothetical prote
16	111.5	5.3	686	1 A59271	trypsin-like prote
17	111	5.3	246	1 DBHU	Ra-reactive factor
18	110	5.2	269	2 C26823	complement factor
19	109	5.2	761	2 JC5759	pancreatic elastasi
20	107.5	5.1	405	2 T35117	brain-specific ser
21	106	5.0	236	2 A28566	probable secreted
22	106	5.0	271	1 ELRT2	T-cell suppressor
23	104.5	4.9	273	2 E85765	pancreatic elastasi
24	104.5	4.9	273	2 H64915	hypothetical prote
25	104.5	4.9	1582	2 T15308	hypothetical prote
26	102.5	4.9	2055	2 T31617	hypothetical prote
27	102	4.8	786	1 A47547	hypothetical prote
28	101.5	4.8	583	2 A29134	serine proteinase
29	101.5	4.8	747	2 I51579	complement factor

30 101 4.8 1047 2 A55617 masquerade precurs
31 99 4.7 274 2 S40004 trypsin-related pr
32 97.5 4.6 259 2 S68424 allergen Der f II
33 97.5 4.6 272 2 JC4170 trypsin-like prote
34 96 4.5 407 1 KPE07 trypsin-like prote
35 94.5 4.5 409 2 T35118 coagulation factor
36 94.5 4.5 416 1 KPE0 coagulation factor
37 94.5 4.5 1238 2 T34929 hypothetical prote
38 94.5 4.5 2145 2 JC4747 adenylylate cyclase
39 94 4.5 266 2 JC4850 trypsin-like prote
40 94 4.5 624 2 T02289 probable polygalac
41 93.5 4.4 452 1 A30351 coagulation factor
42 93.5 4.4 548 2 D82175 probable trypsin V
43 93 4.4 268 2 S68825 pancreatic elastasi
44 93 4.4 430 1 A24702 serine proteinase
45 92 4.4 260 2 S26043 chymase (EC 3.4.21

ALIGNMENTS

RESULT 1

A45134

endopeptidase (EC 3.4.-.-), glutamate-specific - Bacillus licheniformis

C:Species: Bacillus licheniformis

C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999

C:Accession: A45134; S23078

R:Kakudo, S.; Kikuchi, N.; Kitadokoro, K.; Fujiwara, T.; Nakamura, E.; Okamoto, H.; S

J. Biol. Chem. 267, 23782-23788, 1992

A:Title: Purification, characterization, cloning, and expression of a glutamic acid-s

A:Reference number: A45134; MUID:93054737

A:Accession: A45134

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-316 <KAK>

A:Cross-references: GB:D10060; MID:g216263; PIDN:BAA00949.1; PID:dl001415; PID:g21626

A:Experimental source: ATCC 14580

A:Note: sequence extracted from NCBI backbone (NCBIN:118784, NCBIP:118785)

R:Svensden, I.; Bredam, K.

Eur. J. Biochem. 204, 165-171, 1992

A:Title: Isolation and amino acid sequence of a glutamic acid specific endopeptidase.

A:Reference number: S23078; MUID:92155199

A:Accession: S23078

A:Status: preliminary

A:Molecule type: protein

A:Residues: 95-316 <SVE>

C:Keywords: hydrolase

Query Match 8.8%; Score 185; DB 2; Length 316;
Best Local Similarity 23.2%; Pred. No. 2.9e-08;
Matches 77; Conservative 46; Mismatches 139; Indels 70; Gaps 16;

QY 63 SCGQCQCHKGTPPL----PTVKEAKQYLSYETLYANGSRTEQVGIYLLSSSGGAXXRDSG 119

DB 28 AQAPSPH--TPVSSDPSTK-AETSVYDP-----NIKSDQGLXSKAFTGTG---KVNE 76

QY 120 SSGKSRKK-----ROIYGYDSRFISFGKDFLLNYPFSTSVKLST---GCTGTLVAE 167

DB 77 TKEAEKSPAKAPYSIKSVIGSDDRTRVTN---TTAYPYRAIVHLSSIGSGCTGMNIGP 133

QY 168 XHVLTAACHIDGKT--YKGTQKLRVGLKPKFKDGGGRGANDSTAMPQMKFQWIRVRK 226

DB 134 KTVATAGHCIIYDSSGSFAGTATVSPG-----RNGTS-----YPGSVKSTR 175

QY 227 THVPKGTGNANDIGMDYDVALLELKKPKHKFKMKGIVSPPAKOLPGRIHFSGYDNR 286

DB 176 YFIPSGWRSGNTN-----YDYGAIELSEPIGNTVGVFGYSYTTSSLVGTTVTISGYPGDK 230

QY 287 FGNLYVRCD--VKDETDLVLLYQQDAQPGASGYGVYVPMKROQKWKERKII----GIF 340

DB 231 TAGTQWHSGPITAISETYKLOQYAM-DTYGGQSGSPVFQSSSRNTNCSPGCSLAVHTNGVY 289

QY 341 SGNHWDNGSPQDFNVAVRITPLKYAICYW 372
Db 290 GG-----SSYNGRTTRITREVDNLTNW 311

RESULT 2
A35122
metalloprotease (EC 3.4.-.-) mpr precursor, extracellular - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 27-Jul-1990 #sequence_revision 27-Jul-1990 #text_change 15-Oct-1999
C:Accession: A35122; I40010; A69660
R:Slona, A.; Rudolph, C.F.; Rufo Jr., G.A.; Sullivan, B.J.; Theriault, K.A.; Ally, D.; F
J. Bacteriol. 172, 1024-1029, 1990
A:Title: Gene encoding a novel extracellular metalloprotease in Bacillus subtilis.
A:Reference number: A35122; MUID:90130256
A:Accession: A35122
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-313 <SLO>
A:Cross-references: GB:LI0505; NID:gl43209; PIDN:AAA22604.1; PID:gl43210; GB:M29036
R:Smith, R.; de Jong, A.; Bron, S.; Venema, G.
Gene 70, 351-361, 1988
A:Title: Characterization of signal-sequence-coding regions selected from the Bacillus s
A:Reference number: I39994; MUID:89108019
A:Accession: I40010
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-60, 65, 'L', 67, 'S', 69, 'AQA' <RES>
A:Cross-references: GB:M22916; NID:gl43701; PIDN:AAA22832.1; PID:gl43702
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Bruschini, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Ferrari, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y, M.; Ogawa, K.; Ogihara, A.; Oudeg, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serod
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Toononi, A.; Tosato, V.; Uchiyama
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A. Bacillus subtilis.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033
A:Accession: A69660
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-313 <KUN>
A:Cross-references: GB:299105; GB:AL009126; NID:g2632457; PIDN:CAB12018.1; PID:ell182176;
A:Experimental source: strain 168
C:Genetics:
C:Keywords: hydrolase

Query Match 6.2%; Score 130.5; DB 2; Length 313;
Best Local Similarity 22.6%; Pred. No. 0.0013;
Matches 87; Conservative 38; Mismatches 139; Indels 121; Gaps 19;

QY 28 WKPTWPAIRLPVLPQSTLNLRPFDAEAKLEVS-----SSCGPQCHKGTPLPTTYKEAK 82
Db 7 FRKWFAYLTVLCIALA-----AAVSGVPAKAAENPOTSVSNTGKEA-DRTKNOTSKADQ 61

QY 83 QYLSYE-----TLYANGSRTEXQVGIYILSSGGDGAAXRDSGSSKSRKRQIYGYDS 135
Db 62 VSAPYEGTGTSKSLY--GGQTELEKNIQTLQPS-----SIIGTDE 100

QY 136 RFSIFGKDFLLNYPFSTVSKLST-----GCTGTLVAEXHVLTAACHIH----- 178
Db 101 RTRI---SSTTSFPYRATVQLSITKYPNTSSTYTGCTGLPNPNTVVTAGHCVCVSODHGWS 157
QY 179 -----DGKTYVKQKLRVGLFKPKFKDGGGRANDSTSAMPEQMKFQWRKRVHP 230

Db 158 TITAAPGRNGSSYPYGY-----SCTMFYSVK-GWTESKDNYD 195
QY 231 KGWIKGN---ANDIGMDYDVALLELKKPKKFKMGVSPPAKQLPGRITHFSGYDNDRP 287
Db 196 YGAIKLNGSPGNTVGW-YGYRTNSSP-----VGLSSVTGFPCKDTFGTWSDTKP 247
QY 288 GNLVRECDYKDETYDLYQQCDAQPGASGYGVYVVMKQKQKWKRIIGFSGHOMVD 347
Db 248 IR-----SAETYKLTLY-TTDTYGCQSGSPVY-----RNYSDTGQTATAIHT----- 287
QY 348 MNGSPQDFNVAVRITPLKYAICYW 372
Db 288 -NGG-SSYNLGTRVTDVFNNTQIW 310

RESULT 3
EXRT
coagulation factor Xa (EC 3.4.21.6) precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Jan-1995 #sequence_revision 07-Feb-1997 #text_change 08-Dec-2000
C:Accession: S49075; JC4670; PS0191; PS0190; I62745
R:Stanton, C.; Ross, P.; Hutson, S.; Wallin, R.
Thromb. Res. 80, 63-73, 1995
A:Title: Evidence for competition between vitamin K-dependent clotting factors for in
A:Reference number: A58498; MUID:96093366
A:Accession: S49075
A:Molecule type: mRNA
A:Residues: 1-482 <STAI>
A:Cross-references: EMBL:X79807; NID:g506600; PIDN:CAA56202.1; PID:g506601
A:Note: Submitted to the EMBL Data Library, June 1994
A:Note: neither the complete nucleic acid sequence nor the complete translation are s
R:Stanton, C.; Ross, R.P.; Hutson, S.; Wallin, R.
Gene 169, 269-273, 1996
A:Title: Processing and expression of rat and human clotting factor-X-encoding cDNAs.
A:Reference number: JC4670; MUID:96194815
A:Accession: JC4670
A:Molecule type: mRNA
A:Residues: 1-482 <STA2>
A:Cross-references: EMBL:X79807; NID:g506600; PIDN:CAA56202.1; PID:g506601
A:Experimental source: Cos-1 cell
R:Enjoji, K.; Miyazaki, K.; Kato, H.
J. Biochem. 109, 890-898, 1991
A:Title: Characterization of rat factors X and Xa: demonstration of factor Xa in rat
A:Reference number: PS0190; MUID:92041742
A:Accession: PS0191
A:Molecule type: protein
A:Residues: 41-58, 'X', 60-65 <ENJ1>
A:Accession: PS0190
A:Molecule type: protein
A:Residues: 183-186, 'X', 188-207 <ENJ2>
R:Murakawa, M.; Okamura, T.; Kamura, T.; Kuroiwa, M.; Harada, M.; Niho, Y.
Eur. J. Haematol. 52, 162-168, 1994
A:Title: Analysis of the partial nucleotide sequences and deduced primary structures
A:Reference number: I46196; MUID:94222160
A:Accession: I62745
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 295-383, 'G', 385-455 <MUR>
A:Cross-references: GB:D21215; NID:g415309; PIDN:BAA04756.1; PID:g455396
C:Function:
A:Description: catalyzes the proteolytic activation of prothrombin to thrombin in the
A:Pathway: blood coagulation
C:Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homol
C:Keywords: beta-hydroxyaspartic acid; blood coagulation; calcium binding; carboxyglu
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-40/Domain: propeptide #status predicted <PRO>
F:45-84/Domain: Gla domain homology <GLA>
F:41-179/Product: coagulation factor X light chain #status predicted <LCH>
F:90-121/Domain: EGF homology <EG1>
F:129-164/Domain: EGF homology <EG2>
F:183-482/Product: coagulation factor X heavy chain #status predicted <HCH>
F:183-231/Domain: activation peptide #status predicted <APT>